nature

NATURE VOL. 308 15 MARCH 1984

Next crucial week for Europe

The European summit next week in Brussels is billed as critical, and the participants say they will not budge on the urgent questions. They should talk less about food and more about trade.

HOPES have been fading fast, these past few days, that the European summit meeting next week in Brussels will solve the problems that defeated its predecessor, last December at Athens. It is asking too much of flesh and blood, of which even the political heads of governments are made, to expect that in a brief meeting they can find solutions that have eluded lesser fry and government officials for the past three months. So the almost certain outcome of next week's meeting is that the outstanding European problems will yet again be shelved. The best chance, for Europeans, is that the shelving will be intelligently done.

There are three immediate sets of European problems, as follows:

- **Agriculture.** The Community is committed by the Treaty of Rome to the support of European agriculture, but this policy has become uncontrollable. The explanation is simple. For the principal commodities produced by European farmers, the European Commission acts as the purchaser of last resort, buying up what farmers cannot otherwise sell at "intervention prices" fixed so as to correspond to the production costs of marginal producers. This is why the Community staggers unpredictably from one surplus to another — butter one year, perhaps wine the next. Farmers have no incentive not to produce even when they know their products will be unsaleable. The farmers at the margin make a poor living, especially because the European Commission has managed to resist the demand that its prices should be linked with the general level of earnings on the grounds that farmers must be as prosperous as people who work in manufacturing industry, but the flight from the land has not been a stampede because jobs elsewhere are so few, suggesting that the agricultural problem might partly have solved itself if Europe had become more prosperous.
- **Inequity.** Because of the agricultural muddle, members of the Community whose products are chiefly manufactured goods make disproportionately large contributions to the Community. Indeed, only Britain and West Germany are net contributors, although the British contribution is skewed by residual imports of foods from British Commonwealth countries. Four years ago, the British Government made such a fuss that it won a handsome rebate from the Community, but that is now in doubt.
- **Money.** Because the cost of buying food that people want is out of control, the European Communities are fast running out of cash. They may be unable to meet their obligations some time in the summer. But farmers are not alone to blame. The Commission's budget, made up of contributions of member governments, is nominally the equivalent of one per cent value-added tax, Europe's own system of taxation once regarded as a rational alternative to company taxation, now a kind of universal sales tax. The member states could authorize an increased levy, but Britain and West Germany are disinclined to do so when the result would merely be larger food surpluses.

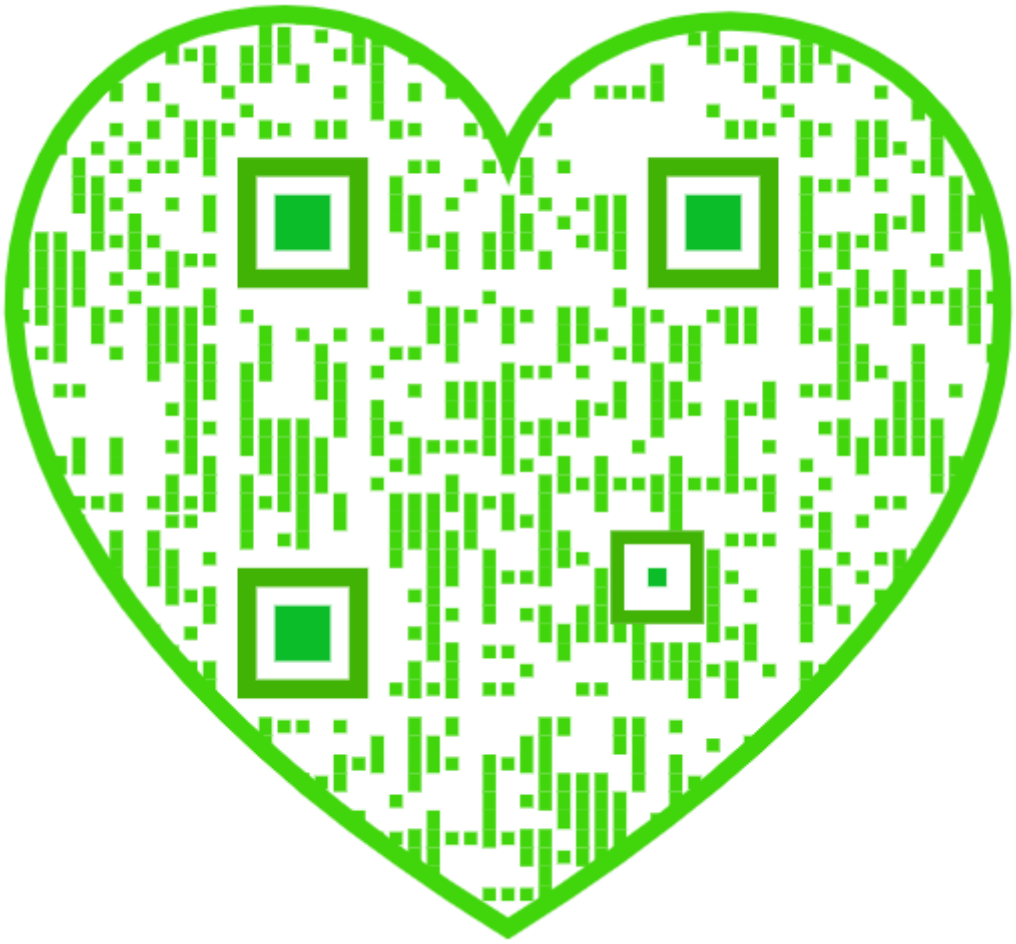
These largely administrative problems, urgent though they are, cannot be quickly solved. The only long-term solution of the problem of surpluses, for example, is that there should be less production, probably by fewer farmers. Next week's summit meeting will not dare admit this openly. Even the British and West German Governments have strong farming lobbies all too ready to fill the streets of the capital cities with whatever seasonal crop happens to be available — artichokes or, worse, pigs — if agricultural interests are threatened. What could be agreed is that intervention prices would not henceforth be increased for commodities that are in surplus, giving high-cost farmers warning to grow other crops or to get out of their declining business.

By itself, this modest step towards good sense would not — and should not — persuade governments to contribute more to Brussels. The justification for that, and the true reason why the immediate problems are so hard to solve, is that the European Community has not been functioning as intended. Non-Europeans are rightly perplexed to know why, more than a quarter of a century after tariff barriers were formally abolished, traffic last month in Europe should have been snarled for more than a week because of the long-standing dispute between international truck drivers and the Italian customs. Why are there still customs posts? Principally to collect value-added taxes and the host of other excise duties that European governments levy on commodities as different as alcoholic drinks and automobiles. The continued existence of these customs posts, which incidentally provide the means by which European governments discriminate administratively against imports and which make the phrase "common market" a misnomer, is an embodiment of the doctrine that even for the purposes of intra-European trade, members states are sovereign and separate. Goods exported from one to the other do not attract value-added tax where they are made but only at the eventual importer's frontier.

If the participants in next week's summit are as serious as they say about seeking a long-term solution for European problems, they must face the simple fact that Europe will not be an integral economic unit while these anomalies persist. True, if trade between Community members were to be regarded not as export-import but as domestic business, governments would have to agree on some common level of value-added tax, and perhaps even on other excise duties, but the benefits would include a shift of revenues away from governments whose farmers produce too much food (which does not attract value-added tax). Better, the Commissions's revenue would be increased, both directly and by the general increase of economic activity that would follow. Nobody pretends that, by itself, such a radical change would be any easier to negotiate than the overdue reform of the European agricultural policy. But skilled negotiators know that it is often easier to deal with two complicated issues at the same time because people who cannot budge on one issue will compromise for concessions on another.

The obvious objection to this or any other package designed to help solve the immediate European problems (central help with national welfare costs is another candidate) is that it would necessarily entail some erosion of national sovereignty. Next week's summitteers should be ready to say "So what?" firmly to themselves before putting this argument into words. For in the mess in which they find themselves, only a further decisive step towards commonality offers an escape. Moreover, unless there is a serious attempt to make the European Community into a common market (with what that implies for impartiality in public purchasing as well), brave schemes such as the Esprit programme of research and development on information technology will turn out to have been pointless. So, since abandoning the Community would make an even bigger mess for everybody, next week's summit has no choice but to grasp two nettles, not just one. □

© 1984 Nature Publishing Group



Neglect of research

British policy on higher education has damaged the research enterprise, perhaps irreparably.

NEARLY three years have now passed since the British Government, in the person of Sir Keith Joseph, still the Secretary of State for Education and Science, insisted in a letter to the University Grants Committee that the government, which meets most of the cost of higher education, must, by extension, also have a right to say how the system operates. The logic is impeccable, and is not in dispute. But it is alarming that the British Government should have embarked on the exercise of its rights with such a narrow view of the purpose of higher education as that put forward last week in Sir Keith's paper to the National Economic Development Council (see p. 217). The minister's insensitivity on this point is a sufficient explanation of his government's indifference to the damage being done to British higher education.

Take, for example, research. Sir Keith noted last week that the government has maintained the value of the research council budgets (which is formally true) and that university research is jointly supported by the research councils and by the recurrent grants that universities receive on the recommendation of the University Grants Committee (which is no longer true, because universities can no longer pay their share) and went on to promise even more vigorous pressure to encourage higher education to meet employers' needs, for graduates and for understanding. The message seems to be — and Sir Keith has only himself to blame if he is misunderstood — that in the months ahead, civil servants will be exercising their ingenuity to open the tap marked "applied research and industrial collaboration" and, because "resources are finite", to close that marked simply "research". The flaw in this ambition is the supposition that research, like water, is a commodity whose supply can be regulated externally at will and whose character can similarly be determined. The British Government seems not to know that the damage already done to the research enterprise in Britain has compromised the capacity of the system to carry out the useful tasks Sir Keith is expecting of it.

Briefly, and necessarily qualitatively, over the past decade there has been a steady decline of the relative quality of the British research enterprise. So much is obvious to those who travel abroad, to the United States in particular, but can also be inferred from the interest of the articles appearing in the journals. The sharpest difference between Britain and the outside world is structural. In the United States, it is plainly still possible for young men and women to gather small teams of colleagues for the pursuit of specific objectives, and to hope that continuity over several years will be provided by the successful competition for grants. By contrast, the British device of scattering (by September 1985) roughly a thousand newly appointed lecturers throughout a university system depleted by the early retirement of senior and not so senior people is an entirely inadequate means of arranging that new research projects will be prosecuted imaginatively and vigorously. The result is that the British research enterprise has come to lack sparkle; too much is imitative, even derivative, too much is not followed through, too little is published with elan. And the condition of university research is faithfully mirrored, even exaggerated, by what happens in government laboratories responsible for basic research. Many once adventurous laboratories have become places where ageing investigators brood on their pension rights.

No great acumen is needed to account for this state of affairs. The claim that the "science budget" has been protected cuts no ice while general support for university research has declined. During the past four years, by contrast, the funds available in the United States for basic research have grown rapidly. During this same period, in Britain, the external demands on research budgets have grown quickly, not merely because of commitments to international research organizations but also in the search for industrially relevant projects. Meanwhile, those responsible for the grant-making machinery have become cautious and over-concerned with the pursuit of the appearance of fairness, which means equal misery for all, while academic researchers have become literally distracted by conflicting demands on their time and energy — teaching, consultancy, meetings to decide the future (if any) of the institutions to which they belong. It is just within the bounds of possibility that Sir Keith Joseph may be able to conjure from this sad and unexpected mess the engine of industrial innovation he is seeking. But it is just as likely that, like the dog in Aesop's fable, he will have let go the bone he is carrying for the sake of its reflection in the water. Certainly he is likely to be the last in his position able with false modesty to boast that Britain is flush with new ideas but dilatory at their exploitation. And undermining the quality of basic research will soon mean that applied research is also worth little. □

Eugenics in Singapore

Mr Lee Kuan Yew should drop his naive eugenic plan to increase the IQ of his community.

AFTER the economic miracle that has made Singapore in the past twenty years, Mr Lee Kuan Yew, the Prime Minister, seems now to be engineering a eugenic muddle. That is the almost certain outcome of the scheme announced last week by which it is hoped that university graduates will be induced to have more children than non-graduates by the offer of freedom to choose the schools their children attend. Even if the underlying assumption that intelligence is largely inheritable were correct, the prime minister seems not to have calculated as intelligently as he might what the social and economic consequences will be. The irony is that Singapore's staggering economic achievements, in their way more remarkable than those of Japan because they turn less directly on manufacture, appear to have been the result of free access to a good but selective educational system.

For the sake of argument, suppose that intelligence is indeed largely inherited in the sense that the chance that the offspring of parents in the top 10 per cent of the IQ distribution will themselves be in that top 10 per cent is, say, x, a positive number less than unity. For convenience, assume that marriage is rigidly stratified by IQ and that, in the absence of special incentives such as those now proposed, the IQ of the Singapore population is unchanged from one generation to the next. If the plan in Singapore is to establish a privileged educational system for the children of parents with high IQ, their fertility will have to be increased by some factor y (a positive number) such that y is greater than $(1/x)$ if the number of educated high-IQ adults is to be increased. If x were as great as 0.8 (the most that even the most zealous hereditarians would claim), the fertility of high-IQ couples would have to increase by a quarter just to maintain the present supply. If, allowing for the notorious lack of perfect correlation between university entrance and IQ, and the imperfections of IQ as a measure of intelligence, x is more likely to be 0.5 or thereabouts, graduate couples will have to double their fertility simply so that Singapore can stand still. In his own defence, of course, Mr Lee Kuan Yew may say that he does not intend that the bright children of non-graduates should in future be denied a decent education, in which case the value of his incentive will be diminished. Either way, he cannot seriously have considered the implications of what he is about.

The underlying flaw, in this and other attempts to exploit the supposed (and probable) heritability of IQ for social purposes, however, is that the side-effects invariably dominate those intended. Does Singapore, which has spectacularly thrived by the exploitation of free-flowing upward mobility, now seriously want to saddle itself with an educational system based on privilege, and on the social rigidity that must follow? And can Mr Lee and his colleagues in the Singapore Government seriously believe that in the future they can educationally short-change the intelligent offspring of less intelligent parents without inviting social unrest of a kind that Singapore has managed to avoid these past twenty years? Mr Lee has been lucky so far, but his amateurish eugenic experiment could change all that. □

© 1984 Nature Publishing Group

Unesco

Still hopes for reversal of US decision

Washington

THE United States may after all stay in the United Nations Educational, Scientific and Cultural Organization (Unesco) after the end of the year, when it has said it will leave unless major changes are made in the agency's management and ideology. Congress, which was not consulted by the State Department before the US withdrawal was announced last December, has begun to take a lively interest in Unesco. Congressional accountants will shortly arrive in Paris to take up an offer to look at the agency's books. And Assistant Secretary of State Gregory Newell admitted last week that it would "intellectually dishonest" to declare the US decision final until its delegation actually leaves Unesco on 31 December.

The State Department has reluctantly promised to give Congress whatever help it needs, but it does not expect the congressional inquiry to find important new information. Mr Newell said the review already conducted by the executive branch had been extensive; 13 government agencies and departments and 83 overseas embassies had taken part. The congressional initiative, led by New York Democrat James Scheuer, will be a comparatively modest affair restricted to an examination of financial management procedures.

Yet the congressional investigation, coupled with the start of hearings last week by the House of Representatives Committee on Science and Technology, has encouraged speculation that the US pull-out may yet be averted. The renewed debate has also encouraged the State Department to reveal more findings of its own review.

This newly-released information confirms that the decision to withdraw was taken despite unambiguous warnings from the US scientific community that the weight of tangible benefits "clearly justified" continued membership. A survey conducted for the State Department by the National Science Foundation (NSF) concluded that Unesco gives US scientists valuable access to research resources and data, enables the United States to share the costs of large international projects and provides contact with scientists with whom Americans might not otherwise work.

The NSF survey also acknowledged that the agency is imperfect, that its scientific effort is too diffuse, that administrative costs are high and that the quality of scientific staff recruited from developing countries is often poor. But it suggested that the United States itself is at least partly to blame. The absence of a central body to coordinate US participation has made it difficult to persuade eminent scientists to join Unesco projects. And the effectiveness of the US Mission in Paris has been impaired by cuts in its staff and budget.

The specific benefits of continued membership, pinpointed in NSF's submission to the State Department, included:

- Participation in a government-level body to attack problems of global significance such as the ecological research conducted through the Man in the Biosphere (MAB) programme, the geological research of the International Geological Correlation Programme (IGCP) and earthquake research conducted by the Natural Hazards Programme (NHP).
- Access to research data, particularly marine data.
- Cost-sharing and access to international facilities, such as the International Centre for Theoretical Physics at Trieste.
- Opportunities to work with scientists from countries with which the United States otherwise has little contact.

Withdrawal from Unesco, the report concluded, would weaken the United States' scientific leadership in the world and contribute to the further politicization of the agency. The State Department's minimum conditions for reviewing its decision to leave, according to Mr Newell, are an end to what the State Department regards as an already intolerable degree of politicization, a budget freeze and abandonment of the agency's controversial new world information and communications order. He said the United States does not expect those changes to occur.

Congressional officials complain privately that the State Department does not really want the agency to reform itself because the Reagan Administration sees withdrawal from Unesco as a useful warning to the United Nations as a whole. Congress, on the other hand, genuinely wants to see Unesco mend its ways, but believes the removal of its director general, Mr Amadou-Mahtar M'Bow, must be achieved first. The State Department, in contrast, stresses that it has no personal quarrel with Mr M'Bow, except for his recruitment to the secretariat of individuals "who have served him poorly". Meanwhile, it is embarking on a vigorous diplomatic campaign to explain its reasons for leaving.

The department claims that in consultations with 115 countries which are members of Unesco there has so far been "near-unanimous" agreement with the findings of the United States' review of Unesco, and some 15 countries are now said to be undertaking similar reviews. The meetings are ostensibly designed to explain its decision to foreign countries and to ensure that the United States' final year in Unesco is a productive one. But they serve equally well to keep the agency guessing about what the United States really intends to do at the end of the year.

Peter David

Los Angeles reactor

Planned scramble for Olympics

Los Angeles

BOWING to pressure from state politicians, the University of California at Los Angeles (UCLA) last week agreed to shutdown its research nuclear reactor during the Olympic Games to guard against the possibility that terrorists might attack the facility. UCLA will also erect concrete barriers around the building housing the 100-kilowatt reactor and post 24-hour security guards outside.

The measures were requested last month by state Assemblymen Gray Davis and Mike Roos, both of whom represent communities surrounding the university. "There are a number of eminent scientists who feel the reactor poses a threat to the community and is an inviting target to terrorists", Mr Davis said. Local citizens fear the highly enriched uranium could be stolen and made into a bomb or that the fuel could somehow be dispersed.

The university is cooperating with the demands but feels the dangers are greatly exaggerated. According to Walter Wegst, UCLA's director of research and occupational safety, the reactor is completely surrounded by tons of concrete and slabs of graphite. "Even if you got into the facility", he said, "it would take a day or more, maybe two days, to reach the fuel." The Argonaut-type reactor is is housed on the ground floor of an eight-storey building. Any explosion large enough to damage the reactor would destroy the entire building, Mr Wegst said, causing it to collapse on the reactor, sealing in the fuel and dangerous radioactive isotopes.

"We think the fuel would probably be more protected from the environment after the building collapsed on it that it is right now", Wegst said. To "play it very safe" and protect the people who work in the building, UCLA will go along with the extraordinary measures, a campus spokesman said.

The UCLA research reactor is still subject to relicensing hearings before the Nuclear Regulatory Commission which have been going on since 1980. At issue is the question of whether small research reactors should use highly enriched uranium fuel — which could attract terrorists — or switch when feasible to low enriched uranium fuel. The conversion to low-grade fuel is under study at UCLA but the material is not commercially available.

Sandra Blakeslee

© 1984 Nature Publishing Group

US space station

Japan reassured on military uses

Tokyo

A MAJOR obstacle to Japanese participation in the US space station project seems to have been overcome this week with assurances from Mr James Beggs, administrator of the US National Aeronautics and Space Administration (NASA), that the station would not be put to military use.

Mr Beggs arrived in Tokyo on Sunday to offer the Japanese Government a chance to join in the project to put a station, permanently manned by a crew of seven or eight, into orbit by the early 1990s. Discussions with Prime Minister Nakasone and the Space Activities Commission — the highest space policy body in Japan — took place on Monday. Because the Japanese constitution specifically forbids the maintenance of armed forces, it would be almost impossible for the government to make a commitment to a space station that might be used for military purposes. But Beggs stated that the US military has no desire at all to participate in the development of the station. It would, he said be launched into a low-inclination orbit (28.5 degrees, 300 km elevation) that would not take it over any of the high-latitude areas which might be of interest to the military. Should the military later want to use space station technology for their own high inclination orbit station, they would not have an automatic right to use technology developed by Japan for the station. Rather, Japan's "intellectual property" would be protected in a way satisfactory to Japan. But Beggs hoped that all work done on the project would "in the NASA tradition" eventually be published and made common property.

With the Japanese Government apparently satisfied on that score, the pressure is on the Japanese to decide quickly to what extent they are willing to participate financially.

Although an official of the Science and Technology Agency said before Beggs' arrival that it was unlikely that a decision could be made before the end of the year, Beggs made it clear that the project will be starting to move rapidly by the last quarter of this year. NASA would like to see the station operational by 1992 and is preparing to invest $8,000 million in the initial development period. By the spring of 1985, some $30 million will have been spent and a decision made on the basic form of the station. Expenditure will then rise rapidly to $150 million by 1985–86 and $400–$500 million by 1986–87.

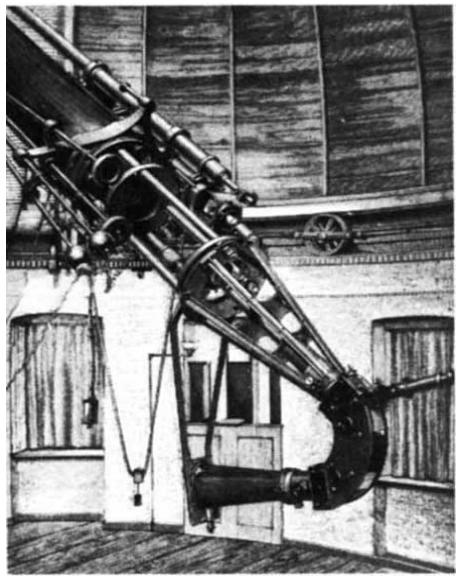
If Japan is "not well on board by the end of this year", Beggs said, "then doors will start to close". This deadline is particularly important because Japanese engineers hope to design totally one of the modules that would be clustered together to make the station. For this, Japan must join in early so that the module interface requirements can be decided.

The United States would like Japan to contribute an amount equal to 10–15 per cent of the US investment, to be added to the 20–25 per cent asked of Europe and Canada. If all countries participate as asked, the development budget will be near the $11,500 million dollar mark, allowing a far more sophisticated programme than the United States could accomplish on its own. Beggs stressed that the money spent should be viewed as an investment: space projects, he said, can return substantial sums both in direct returns on satellite launches and in high technology spin-offs, and the aim is to make the space station a commercial and industrial success.

But with the Japanese Government already struggling to find funds for its own ambitious space programme (*Nature* 1 March, p.3), it is hard to see where the money will come from. Beggs said that the United States is not insisting on a 10–15 per cent investment from Japan — each nation will have to decide for itself how much it wants to invest. All Michiyuki Isurugi, director-general of the Science and Technology Agency, is prepared to say so far is that the meeting with Beggs has "deepened interest in the project", but discussions of more concrete plans have not yet begun. **Alun Anderson**

Phaenomenal colours

THE early history of the spectroscope since 1666 when Isaac Newton ". . .procured me a Triangular glass-Prisme, to try therewith the celebrated Phaenomena of Colours" is covered in a new exhibition which opened last week in Cambridge, England, and runs until 7 December. The exhibition is at the Whipple Museum of the History of Science, part of the University of Cambridge. A catalogue and illustrated history of the spectroscope are published to accompany the exhibition. The engraving below shows a two-prism spectroscope fitted to an 11-inch refractor at Potsdam Observatory. □



Commercial space

Poor outlook for Landsat sale

Washington

COMPANIES interested in buying all or part of the United States' Landsat satellite remote sensing system have until 19 March to submit their bids to the Department of Commerce. The Communications Satellite Corporation (Comsat), the Radio Corporation of America (RCA), Fairchild, General Electric and a new company called Space America have already expressed interest. But the proposed sale is regarded by many outsiders as an unpropitious start to the commercialization of space.

Part of the problem is financial. What little market exists for remote-sensing data consists mainly of government departments. As the proposal to commercialize the system has creaked through Congress, even enthusiasts for private sector investment in space have questioned the wisdom of the government selling a system and buying back its products. A proposal by the Reagan Administration to include government meteorological satellites in the sale was rejected outright.

But many in Congress, prompted by the Office of Technology Assessment (OTA), fear that sale of the Landsat system will also jeopardize the diplomatic interests of the United States. An OTA technical memorandum published this month says that Landsat has been used over the years to advance important foreign policy objectives — notably preserving the principle of the free international flow of information and helping developing countries see the American use of space as an opportunity rather than a threat.

The commercial interests of a private buyer may well conflict with such objectives, OTA warns. On the other hand, any constraints the Commerce Department intends to place on a sale to ensure that Landsat data remain available to developing countries at an affordable price threaten to make the enterprise unprofitable.

The OTA memorandum points to several other problems raised by the sale. For one thing, the defence and intelligence communities are the biggest users of Landsat data within the federal government. Unless there were safeguards on a privately-owned system, these communities might find it necessary to build and operate their own systems, a course that would probably cost the government more than it would save by selling Landsat.

OTA also questions whether a private owner would be far-sighted enough to develop remote sensing technology at the pace necessary to allow the United States to keep ahead of foreign competitors. The European Space Agency (ESA) and several individual countries have well-advanced plans to launch their own systems and to

© 1984Nature Publishing Group

market the results.

Most worrying for the United States is the French SPOT satellite, due to be launched next year. A remote-sensing satellite capable of high-resolution multispectral stereo-images of the land surface, SPOT is billed by the French as the world's first commercial remote-sensing satellite system. In fact it is heavily subsidized by the French Government.

Competition such as that posed by SPOT puts the Commerce Department in a cleft stick. The Reagan Administration believes in the power of the private sector to maintain US technological leadership, but in the case of Landsat some form of subsidy, either direct or indirect, may be necessary. What is not clear is whether Congress will go along with plans to sell a public asset if it is going to continue to be a burden on the federal budget.

Peter David

UK higher education

More talk at NEDC

THE British Government seems to have found an unusual forum for developing its strategy for the future of higher education — the National Economic Development Council.

Discussion of a paper on higher education prepared for last week's meeting of the council by the Secretary of State for Education and Science, Sir Keith Joseph, seems to have surprised the minister by its length but to have raised in the minds of others the question whether the British Government still needs the University Grants Committee.

The council is the tripartite talking-shop which normally represents government, the Confederation of British Industry and the Trades Union Council, but whose meeting last week was boycotted by the unions, still smarting from the government's decision to quash the rights of intelligence gatherers to join trades unions.

The striking feature of Sir Keith's paper is the evidence it provides of the government's determination to direct development of education from the centre. Its starting point is a declaration that, apart from the contribution of higher education towards "national well-being and society as a whole", British higher education must make "judgements" of the demand for graduates in different specialities and "contribute directly to mutually beneficial collaboration with employers".

Acknowledging the "case against any detailed or comprehensive planning of graduate supply", Sir Keith nevertheless says that the government thinks it right to "give a broad steer to the system to some extent". Its intervention so far has been the "indication" to the universities that science, technology, engineering and "other vocationally relevant forms of study" should recruit more students, the 15 per cent increase of engineering students and 50 per cent increase of mathematics and computing students planned for next year in polytechnics and the provision of 5,000 places on information technology graduate courses planned for 1985-86.

Further ahead, the government hopes for a 50 per cent increase (to 15,000 a year) in the output of graduate engineers in the decade to 1987-88, and the minister's paper says that he is "actively considering" what should be the target for succeeding academic years, when the age group from which students are recruited to higher education will be shrinking.

The paper also seeks to stimulate the job market by noting that British industry must make employment for engineers more attractive, saying that "student demand will quickly react against engineering" if jobs do not materialize. Sir Keith says that there will be more jobs for electronic engineers but fewer for civil engineers, and that physics in higher education will be squeezed by competition from engineering. Acknowledging that engineering courses cost more than others, Sir Keith offers guarded hope of financial relief for higher education with the remark that "resources are finite and increased expenditure must therefore be clearly justified".

Research is described in the minister's paper as "a further area where higher education and employers must come together productively". Sir Keith says that "fundamental research . . . is vital to our scientific progress, but is not necessarily directed at a specific industrial application". For the future, the minister announced that the government would decide how far the switch in higher education towards science, technology and engineering should go.

Employers, for their part, were asked to say what their expectations are of higher education, to describe their plans for employing the new graduates at attractive salaries and to comment on "the value to the economy of basic research financed by the research councils". □

UK universities

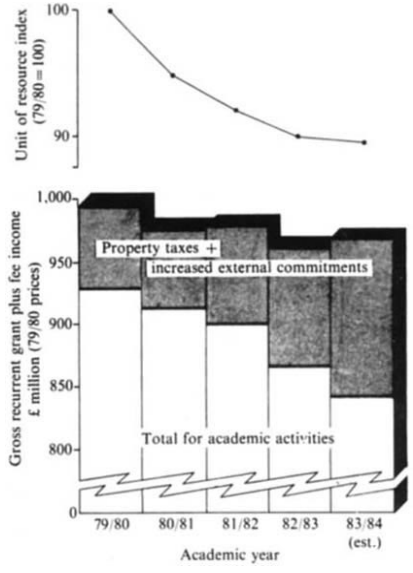
Row about budgets resolved

RECENT trends in financial support for British universities have lately been a matter of public dispute. Mr Peter Brooke, Under Secretary of State with special responsibility for higher education, sent indignant academics diving for their calculators when he claimed in a letter to *The Guardian* newspaper that the universities' recurrent grant had shown a 25 per cent cash increase and "virtually no change in cost terms" between 1980–81 and 1983–84. That view was quickly challenged and, after questions in Parliament, Mr Brooke admitted there had been a real reduction of 8.75 per cent.

Mr Brooke's mistake was that he failed to allow for changed arrangements in 1982–83, when tuition fees were reduced and the recurrent grant channelled through the University Grants Committee was increased to compensate. A better picture of trends as they affect universities is given in the figure above, which covers both sources of income. The deflator is the Treasury-approved Tress–Brown index of universities' recurrent costs.

During the five years since 1979, several additional commitments have been imposed on the universities over which they have no control, thus making comparisons difficult. These include redundancy compensation (£26 million in 1983–84), the upward incremental drift of salary gradings, increased superannuation payments and a shift of funds from capital grant. Together with rates, which the universities never see, these non-academic increases are lumped together in the upper half of the histogram.

The total left over for academic activities compares like with like (although it does include the targeted "new blood" and



information technology schemes in 1983–84). The shortfall of £112 million at 1979–80 prices is a drop of 11.5 per cent in income needed to maintain activities at their 1979 level. Full-time student numbers rose from 293,000 in 1979–80 and then fell back to 288,000 in 1983–84: the resulting "unit of resource index" is also shown above. The Committee of Vice-Chancellors and Principals, which supplied the information, points out that these reductions have been achieved despite a shift towards the more expensive science-based subjects. A drop of 1 or 2 per cent a year in the unit of resource has been suggested, and universities' total recurrent grant next year will be 2.5 per cent lower than was announced in November.

Tim Beardsley

© 1984 Nature Publishing Group

Radioactive waste

British search for dumps opposed

THE British search for radioactive waste disposal sites has run into trouble. Last October, the Nuclear Industry Radioactive Waste Executive (NIREX) flagged its intention to investigate two land sites as permanent dumps for contaminated material from the nuclear industry, classified as waste of low and intermediate activity. Now, Imperial Chemical Industries Limited (ICI), the owner of a disused anhydrite mine at Billingham in north-east England, says that it will oppose the plan to investigate that site, while local councillors say they have evidence that the second site, at Elstow in the Bedfordshire brickfields of central England, is unsuitable for its intended purpose.

ICI's anhydrite mine has been sealed for some years and the on-site investigations NIREX plans would probably require planning permission. ICI said this week that it does not intend applying for permission to open the mine, since it has no commercial interest and prefers to maintain good relations with the local community. A spokesman said it had become clear since October that local opinion is intensely hostile. When NIREX first announced its choice of the site as a possible permanent resting place for long-lived intermediate level wastes, it coyly did not mention that military wastes might also end up there. The outcry when that emerged some weeks later probably made NIREX wish it had been more forthright.

Even so, NIREX intends to press ahead with its Billingham plan without ICI's co-operation, although both sides admit that the legal position is far from clear. Because NIREX has no powers of compulsory purchase, it might have to face the rigours of a full public inquiry about Billingham armed only with the data it already has. ICI makes no bones of the fact that it would actively oppose NIREX at a public inquiry, on the grounds that good community relations are essential for the sake of its own business.

At Elstow, if the site proves suitable, NIREX plans to excavate a series of trenches in the local bed of Oxford clay to take shorter-lived wastes. It has now become clear that the clay bed is only 15–20 metres thick, although rather deeper layers were previously mentioned. The base of the clay layer is uneven and local hydrogeologists say there may be a fault on the site: there are certainly faults in the clay layer within 2 km. There is also concern that nearby excavations where clay has been extracted in the past may make it impossible to predict groundwater flow with sufficient accuracy. NIREX will have to show that radionuclides will not within 200 or 300 years percolate through their encasing concrete blocks and enter the biosphere in significant amounts. Some of the nearby excavations are filled in with domestic refuse, again leading to uncertainty about patterns of groundwater flow.

A further complication is that the London Brick Company has permission to extract clay from the field immediately adjacent to the proposed site and has applied for planning permission to excavate the proposed dump site itself. There is little doubt the local authorities would agree to that request if it offered a way of keeping NIREX out, but the Central Electricity Generating Board, which owns the site and is an industrial partners in NIREX, is unlikely to give away one of the few possible sites for a dump over which NIREX already has control.

The site is well served by transport links and is already something of an eyesore, so countryside conservationists are unlikely to object to its use. Inevitably, however, local people accuse NIREX of putting convenience before safety and say a more remote site should have been chosen. NIREX itself remains — or claims to remain — serene, although five local authorities have united to fight its plan. The authorities are making strong representations to government ministers now that it is known that NIREX intends shortly to submit a planning application. NIREX says no site-specific safety criteria have yet been drawn up and that all indications are that the Elstow site is suitable.

Tim Beardsley

Looking north-east over the Elstow site

Soviet nuclear power

Teething trouble persists

THE Soviet nuclear power industry, which is running badly behind schedule, may have been granted a temporary respite by the death, last month, of the party leader, Mr Yurii Andropov. During the Central Committee plenum in June 1983, Mr Andropov gave his strong personal backing to the new energy programme involving the construction, over the next ten years, of 30 nuclear power stations, 12 of which will also provide district heating for cities in the European part of the Soviet Union.

On 8 February, *Pravda* published an editorial attacking delays in construction at many of these sites, including the fast breeder reactor at Beloyarsk. The tone of the article implies that one of Mr Andropov's drives against inefficiency and incompetence was planned for the nuclear construction industry, and that widespread upheavals and dismissals might be expected. Two days later, however, Mr Andropov died, and his successor, Mr Konstantin Chernenko, has so far concerned himself with international rather than domestic issues.

For some time there have been indications in the Soviet press that all is not well with the nuclear power programme. Conventionally, delays are blamed on late and insufficient supplies from the manufacturers of construction materials, equipment and generating machinery. The *Pravda* editorial repeats this line, urging severe economic penalties for breaches of contract by suppliers, with matching incentives for those of proven reliability.

Poor organization partly explains apparently contradictory progress reports. In January 1983, for example, Moscow Radio noted that construction of the nuclear power and district heating station in Gor'kii was two months ahead of schedule, but, two weeks later, reported that the whole project was being held up for lack of the safety-housing, already nearly a year overdue. The defaulting supplier was the Volgatsemmash production association in Tolyatti, but a few months later the most important of all suppliers to the nuclear industry, the Atommash association at Volgodonsk, came in for severe criticism in *Pravda* in terms which hinted that safety regulations were being ignored. A top Communist Party troubleshooter had to fly to Volgodonsk and, a few months later, a new high-level committee was set up to monitor safety in the nuclear industry.

Until recently delays in nuclear construction projects have received relatively little publicity, partly because the original target dates are not always public knowledge and partly because the installations most affected in the past few years have been

© 1984 Nature Publishing Group

additional generating sets at a site already operational. The BN-600 fast breeder reactor at Beloyarsk, for example, was scheduled for completion in 1979 and it in fact went critical (with great publicity) in February 1980. The delays mentioned in *Pravda* refer to additional generating sets at the same site which, in the nature of things, receive less attention.

By blaming the suppliers and manufacturers of ancillary equipment, the *Pravda* leader-writer gives the impression that all would be well with the nuclear power programme if the manufacturers kept to their delivery dates. This, however, is not the whole story. Last November, when the reactor for the third generating set at the Kursk power station went critical, almost three years late, the provincial Party Secretary, V. Ivanov, discussed the delay at some length in *Pravda*, on the page normally devoted to Party affairs.

Ivanov attributed the delay not only to a shortage of construction workers, followed by a sudden influx of 6,000 workers with consequent confusion and disorganized working, but also to basic difficulties with generating equipment. The Ministry of Power and Electrification, he said, had not finally solved one of the principal problems, the design of a standard production-model 1000-MW generating set. As a result, there had been, he said, a constant stream of every kind of correction and alteration between the site and the suppliers, so that the amount of installation work on the set was five times normal.

Vera Rich

Israel

Backers for nuclear-free zone

Rehovot

A LEADING Israeli hawk has paradoxically come out against the building of nuclear weapons here, while a card-carrying dove has argued that Israel should openly join the nuclear weapons club. The hawk is Minister of Science and Development Yuval Ne'eman, who said recently in Jerusalem that it would not be in Israel's interests to base its defence on nuclear deterrence, because the nuclear balance of terror that exists between the superpowers would not work in the Middle East.

Professor Ne'eman is a renowned nuclear physicist and was for many years connected with Israel's defence establishment, so he presumably knows what he is talking about when he says that while Israel has a nuclear infrastructure, it "never crossed the threshold into the nuclear weapons club". He is convinced that a nuclear deterrent, which would not in any case be used unless the military threat against Israel were extremely grave, would not lessen its need to maintain conventional forces. Professor Ne'eman also points out that a nuclear device set off close to Israel's borders in order to destroy massed Arab armies would also destroy much of Israel.

The dove who favours nuclear weapons is Dr Shai Feldman of Tel Aviv University's Jaffee Centre for Strategic Studies. His controversial book, *Israel's Nuclear Deterrence: A Strategy for the 1980s,* bluntly calls on Israeli policymakers to make nuclear weapons the cornerstone of their country's defence. He speaks of an arsenal of 30 to 40 weapons in the 20 to 60 kiloton range.

Unlike Professor Ne'eman, Dr Feldman does not believe that contemporary Israel should embrace the West Bank and the Gaza Strip. This is consistent with his call for a defence policy based on nuclear deterrence because, in his view, Israel could withdraw to borders similar to those it held in 1967, "declaring at the same time that any significant attempt by Arab armies to cross these borders would meet nuclear punishment". He believes that an openly nuclear-based defence policy would permit Israel to cut its huge expenditure on conventional weapons and on the maintenance of a large army, which places an increasing burden on the economy.

This view is disputed by many other observers, including Dr Yair Evron, a colleague of Feldman at Tel Aviv University. Evron declares that "no Israeli policymaker would be willing to rely exclusively on a nuclear deterrent".

Both Feldman and Evron doubt whether, in the long run, Israel will be able to prevent its enemies from acquiring nuclear weapons (as Prime Minister Begin said it would after the bombing of the Iraqi reactor in 1981). If they are right, what can Israel do to prevent a nuclear arms race, and perhaps nuclear war, in the Middle East?

Many foreign observers believe that all would be well if Israel would only agree to sign the Non-Proliferation Treaty (NPT), as have most other Middle Eastern states. But this view is not shared by most Israelis. A physicist at the Weizmann Institute, Shalheveth Freier, points out that Libya openly but vainly shopped for bombs before signing NPT, and, he adds, "it is to be doubted whether this switch in declaratory stance amounts to a change of intent".

What Freier, Ne'eman and others would like to see is an agreement among all the countries of the Middle East for the establishment of a nuclear weapons-free zone in this area, like the one set up in Latin America under the Treaty of Tlatelolco. For while NPT allows the country whose nuclear facilities are being inspected to determine who the inspectors will be and when they will come, the Tlatelolco Treaty calls for inspections at any time and by inspectors from the region itself.

Nechemia Meyers

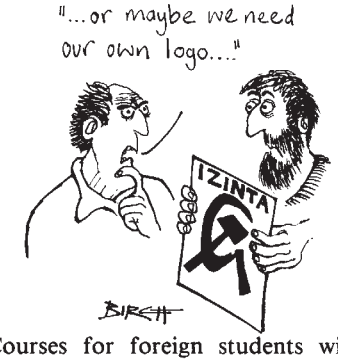
Hungarian universities

Market forces are felt

HUNGARIAN academic science is fighting recent cuts in government budgets by going commercial. The first subsidiary company of a research institute has now been registered, and two Budapest universities have launched plans to admit students from outside the Socialist bloc who will pay their fees in hard currency.

Izinta, the new subsidiary of the Institute of Isotopes of the Hungarian Academy of Sciences, was created to circumvent a government decree reducing the number of people who may be employed by any research institute. The Institute of Isotopes, in addition to wide-ranging programmes of pure and applied research, has had since its foundation in 1959 a production department which supplies the Hungarian market with radiochemicals and irradiation services. Since 1961, the institute has exported its products and services through the state trading company Medimpex.

Hiving off the commercial side of the institute's activities to form Izinta was thus a fairly rational method of reducing staff. Thirty scientists and technicians have been transferred from the institute to the new company, which will return 25 to 40 per cent of its profit to the institute which, in return, guarantees the "productive, technological and economic" basis of Izinta.



Courses for foreign students will be introduced at the Semmelweis Medical University (taught in German) and at the Budapest Technical University, where the languages of tuition will include English.

According to Dr Miklos Ivanyi, vice-rector of the Technical University, a "considerable proportion" of the hard currency thus earned will remain at the disposal of the university, while lecturers teaching the courses may be able to draw part of their salaries in dollars. Since a shortage of hard currency for journal subscriptions, laboratory equipment and foreign travel is a growing problem of Hungarian academic life, other Hungarian universities may well decide to follow suit. Fees, said Dr Ivanyi, have not yet been decided, but will probably be significantly less than at similar establishments outside Hungary.

Vera Rich

© 1984 Nature Publishing Group

Japanese literature

Congress seeks more translation

Washington

US SCIENTISTS are concerned that the flow of technical information between the West and Japan — whose scientific research establishment has now reached third place in the world — is a one-way street. Only a small number of Japanese journals and technical reports are readily available in the West, even in abstract form. Conversely, most Japanese scientists understand written English.

"A very large part of the Japanese output is concealed from us because of the language barrier", says William Budington, librarian of the John Crerar Library in Chicago, an independent research library. Testifying before a House of Representatives subcommittee last week, Budington said that fewer than one-fifth of Japanese scientific and technical journals are surveyed by the major abstracting and indexing services in the West (such as *Chemical Abstracts*). And 75 per cent of the Japanese publications are wholly in Japanese, compared with only 28 per cent of those covered by the Western abstracting services. It is estimated that there are more than 9,000 Japanese scientific and technical journals.

Budington says that there is a demand in the United States for Japanese publications; his library, which receives a few hundred Japanese periodicals, finds them in "fairly heavy" use. He adds that 20–25 per cent of all publications translated from foreign languages that his library receives are from the Japanese. "It's about to outstrip Russian", he says.

Dr Eleanor Westney, acting director of the MIT–Japan Science and Technology Program at Massachusetts Institute of Technology, says the problem is not only that few Americans can read Japanese, but that even fewer can both read Japanese and evaluate technical material. And the lack of information about Japanese research thus engendered feeds on itself: "there is a widespread belief that all important scientific information will be published in English". Dr Westney says that while this may be true for the basic sciences, it is far less so in applied fields. The Institute of Electrical and Electronics Engineers (IEEE) has recently given some credence to that view; its Magnetics Society announced plans to publish complete translations of two Japanese journals on magnetics and a number of special proceedings — a total of 1,500 pages per year. The service, available by subscription, will begin in January 1985. Clark Johnson, president of the Magnetics Society, says that the Japanse counterpart to IEEE publishes 27 journals, many of which would probably also be worth translating.

The programme that Dr Westney directs at MIT is one of only two in the United States aimed at encouraging scientists and engineers to learn Japanese. MIT is sending six students each year to spend one year working in universities, government laboratories and companies in Japan. (Dr Westney, "to lay to rest any lingering suspicious of Japanese 'secretiveness'", says that the MIT programme receives more offers from Japanese companies than it can fill with qualified students.) North Carolina State University runs a similar programme through its "Japan Center", established in 1980 largely to encourage Japanese investment in the state.

Although Dr Westney says that suspicions of "secretiveness" are unfounded, she acknowledges that the Japanese handling of their own literature does make for a certain impenetrability, even to foreigners fluent in Japanese. She says that the Japanese publications are not as well abstracted or indexed, even in Japan, as are Western publications in the West. Further, many researchers in Japan publish in the house organs of their university or company — journals which tend to be deservedly obscure because of their uneven quality. Informal contacts, Dr Westney says, while obviously of importance in the West, become "critical" in Japan if one is to learn about recent important publications.

Representative Doug Walgren (Democrat, Pennsylvania) has introduced in Congress a bill that would allocate $750,000 to the Department of Commerce next year for abstracting, indexing, translating and disseminating Japanese technical information, through either the National Technical Information Service or the Office of Productivity, Technology and Innovation. Some independent efforts in Congress are under way to allocate funds for improved teaching of Japanese, as well as other foreign languages. **Stephen Budiansky**

Applied embryology

No surrogate British mothers

THE use of surrogate mothers to bear children for infertile women now seems unlikely to expand on any organized scale in Britain. The council of the British Medical Association (BMA) last week told physicians that it would be unethical for them to become involved in such schemes, and a private member's bill now being promoted in the House of Commons aims to make it illegal for agencies to offer "womb leasing" services.

The use of surrogate mothers impregnated through artificial insemination by donor has been practised rather clandestinely in Britain, but the improving availability and success rate *in vitro* fertilization is expected to increase the demand for women willing to bear a child for payment of a fee. In the United States, where the practice is more widespread, several states have made it illegal to sell children, but agencies get round such laws by claiming fees as "medical expenses".

In Britain, the government's Warnock committee, set up to advise on the implications of new developments in the field of human fertility, is experiencing difficulties in deciding what it will say on the matter. BMA's decision last week will provide a strong lead: although its recommendations are not binding, they are likely to be accepted by British physicians. The association's central ethical committee received no submissions in favour of encouraging physicians to practise surrogating, and the major national associations concerned with fostering are opposed to its use.

The area is an ethical and legal minefield. Despite the initial appeal of surrogate motherhood as a solution to infertility due to inability to sustain a pregnancy, BMA's ethical committee decided that the probable damage to the child through being born into an unorthodox and perhaps damaging emotional and legal environment should outweigh the desire of the parents for a child.

In Britain the law is not clear on the status of an agreement to bear another's child if the pregnancy is started artificially: in the case of any dispute over such an agreement, it would be for the courts to decide the outcome. The offspring of a surrogate mother is however definitely illegitimate under British law, and it is for the genetic parents to adopt it should they choose to do so. As this entails going through the courts, the affair comes into the public domain.

Dr James Appleyard, who supported last week's recommendation to BMA's council, says the use of surrogate mothers is fraught with difficulties. In Britain the situation has already arisen where a surrogate mother refused to part with the child she had carried, and a court ruled that she had the right to do so despite a previous agreement that the child would be handed over. Even more harrowing scenarios are easily envisaged: what, for example, would be the position if money was paid and then a child was born handicapped?

The parliamentary bill aimed at preventing "womb-leasing" clinics from establishing themselves in Britain may be debated this week but is not being actively supported by the government and so will probably have to be rescheduled. The government, while not opposed to the measures the bill advocates, is inclined instead to await the Warnock committee's recommendations, which are expected in June. But, already, Miracle Programmes Inc. of Washington DC has been approached with a view to opening a surrogate clinic in Britain. **Tim Beardsley**

© 1984Nature Publishing Group

Man or machine in space

SIR — Your leading article on the US manned space programme (*Nature* **307**, 1; 1984) is uninformed and unfair. The view that there can only be manned or unmanned space exploration is a needlessly adversarial position. The National Aeronautics and Space Administration (NASA) actively supports both programmes, which are inexorably linked by spacecraft support and technological advances.

The lack of new unmanned planetary projects is more the consequence of decreased funding than deliberate policy. In the short term, planned unmanned space exploration is dependent on shuttle technology; the Galileo, Mariner Mark II, Observer and Spartan programmes all require it. Moreover, the shuttle and an orbital platform are the key to the future of unmanned planetary exploration. The blanket statement that the shuttle harms planetary science is unfair, shortsighted, and untrue.

So are the statements regarding the historic Apollo programme. The successes of Pioneer, Voyager and Viking depended and continue to depend on technology developed for the Ranger and Surveyor programmes, which were in turn developed for Apollo. Furthermore, the only significant quantities of extraterrestrial materials ever examined by scientists were returned by, as your writer quoted Tom Lehrer, "some clown on the Moon". The statement that Apollo led "nowhere" overlooks at least 16 articles on Apollo-related research published in *Nature* between 1970 and 1972 alone[1].

Your statement that a manned space platform would yield "two decades of original research on why astronauts vomit and whether you can make the perfect ball-bearing in space" demonstrates a lack of familiarity with the findings from Salyut, Skylab and recent shuttle flights. Such a platform will have many uses including satellite repair and maintenance, materials processing, astronomy, biology and basing for unmanned planetary probes.

Your rather shortsighted implication that a manned space platform may be needed in the future, but not right now, smacks of the philosophy of Senator William Proxmire who recently awarded the Golden Fleece Award to NASA, stating that because the planets have been there for "millions of years, and will be around for millions more", there is no hurry to investigate them.

Your leading article underscores a major problem in research planning: short term versus long term. Technology developed for either the manned or unmanned programme ends up in the other, and both projects benefit. Adherents of the two philosophies should not bicker, but should join forces to lobby for increased funds. Most importantly, emphasis should be placed on long-term planning of manned and unmanned space exploration. The shuttle and a permanently manned orbital platform are a good first step.

MARK OWEN MARTIN
Department of Biological Sciences,
Stanford University,
Stanford, California 94306 USA

1. *Nature* **226**, 925 (1970); **228**, 254 (1970); **232**, 94; **234**, 211, 402 (1971); **236**, 215, 381 (1972); **237**, 158, 446 (1972); **238**, 450 (1972); **239**, 205 (1972); **240**, 94, 95, 96, 139, 259 (1972).

Plum pox virus

SIR — In *Nature* (**303**, 365; 1984) under the title "Italy's farmers battle with virus", Robert Walgate described an outbreak of plum pox virus (PPV) around Turin and Bologna. In that article, the phytopathological problem appeared dramatic. I would like to modify the impressions given, and describe the present situation.

The Italian Ministry of Agriculture was informed of the outbreak soon after its discovery, in August 1982. In November 1982, it sent a letter to the regional phytopathological service ordering the destruction of all affected trees. On compensation, the regional administration of Piedmont has agreed to supply healthy trees free of charge to replace those destroyed.

Eradication of infected trees was, however, begun immediately by personnel of the regional phytopathological service. All trees in orchards in the affected area were screened individually for PPV symptoms. The apparently infected plants and those surrounding them were checked for PPV serologically at this institute. The results of this survey will be presented at the "Giornate Fitopatologiche" to be held in Sorrento (Naples) on 29 March 1984. They can be summarized as follows.

PPV was detected only in apricots and in some suckers from wild peach and mirabolan rootstocks of infected apricots, but not in peaches or plums. The outbreak appeared restricted to about 2,500 hectares of cultivated land. During 1982 and 1983 more than 15,000 plants were screened and some 1,400 found to be infected. The two main virus foci detected in 1982 (500 trees) were destroyed the same year while seven other foci detected in 1983 are to be destroyed during the present winter.

The spread of PPV in Piedmont seems to have occurred mostly by vegetative propagation. In only a few cases does the gradient of incidence in the orchards suggest that secondary spread of virus by aphids may have occurred.

These result allow us to hope that the outbreak of PPV may be contained this year. It may even be possible to eradicate PPV from the area within a few years, although in other countries eradication has proved difficult.

MAURIZIO CONTI
Instituto di Fitovirologia applicata,
Consiglio Nazionale delle Ricerche,
Turin, Italy

Dietary cholesterol

SIR — The Lipid Research Clinics Coronary Primary Prevention Trial (LRC-CPPT) showed conclusively that lowering of serum cholesterol by means of a bile acid sequestrating drug reduced the chance of a heart attack. Your editorial[1] suggests that this does not prove the benefit of a reduction in cholesterol intake, because dietary cholesterol has too little influence on blood cholesterol levels in man.

Such a statement combines two errors. First, it equates a cholesterol-lowering diet with a low-cholesterol diet. In man, however, the most important determinant of serum cholesterol is not the amount of cholesterol in the diet but that of saturated fatty acids. Reduction of saturated fatty acid (and cholesterol) intake, combined with an increased consumption of polyunsaturated fatty acids, reduced serum cholesterol by 10 to 15 per cent or more in several long-term mass field trials[2]. This contrasts with 8.5 per cent in LRC-CPPT, which is typical for other drug trials as well[2]. So diet is generally more effective than drugs in lowering cholesterol.

A second error is the suggestion that dietary cholesterol has a negligible effect on serum cholesterol. It is true that if one already eats two eggs a day (equivalent to about 500 mg of cholesterol) one might as well eat four as far as serum cholesterol is concerned, because there is a saturation effect. However, reduction of daily cholesterol intake from the usual US or British intake of about 500 mg to 200 mg will reliably lower serum cholesterol by 4 per cent on average, and by more than that in "hyperresponders" with high sensitivity to dietary cholesterol[3]. Thus the effect of cholesterol in the diet is small, but not unimportant.

MARTIJN B. KATAN
Department of Human Nutrition,
Agricultural University,
De Dreijen 12, 6703 BC Wageningen,
The Netherlands

ANTON C. BEYNEN
Department of Laboratory Animal Science,
State University,
Yalelaan 1, 3508 TD Utrecht,
The Netherlands

1. *Nature* **307**, 205 (1984).
2. Rifkind, B.M., Goor, R. & Schucker, B. *Atherosclerosis* Vol.6 306-310 (Springer, Berlin, 1983).
3. Katan, M.B. & Beynen, A.C. *Lancet* i, 1213 (1983).

Effect and affect

SIR — I should like to draw attention to the recent North American fashion for the use of the word "affect" as a noun. The usage seems to have spread like a forest fire through my American colleagues. I tell them that they can "affect an effect" but not "effect an affect". But they take no notice and persist in this affectation.

R. C. WATKINS
Esso Chemical Research Centre,
PO Box 1, Abingdon,
Oxfordshire OX13 6BB, UK

© 1984Nature Publishing Group

Henge manufacture

SIR — As a mining engineer engaged, like my colleague of some four millennia ago, in the acquisition and carving of stones for a henge, albeit at only half his scale, I have perforce been studying his work.

Two points seem to have been missed in discussion of the original Stonehenge. The first is the mysticism concerning the number 19. G.S. Hawkins has suggested that its relevance was to the period of swing of the Moon, about 18.6 years (*Stonehenge Decoded*, Fontana, London). A simpler reason would be if the Hyperborean month were 19 days. The symbolism of this is that, allowing for 4 feast days, which would be time out of normal work (midwinter, the equinoxes and midsummer), this gives us 19 such months in the year. This fortuituous combination of the 19s would be marked and relate to the 19 bluestones of the inner circle.

The second point deals with transportation. Common opinion suggests that the stones were manually hauled to the site. But surely if oxen were available to provide shoulder blades for shovels, they would also be present to provide haulage for the stones. Given that oxen can pull between 1,200 and 4,000 pounds weight, depending on where they fit in the scale between modern oxen and buffalo, this would reduce the haulage team from 700 men to 15 to 50 animals. In my colleague's defence, I propose this as a much more likely option than that commonly accepted.

DAVID A. SUMMERS
Rock Mechanics and Explosives Research Center, University of Missouri-Rolla, Rolla, Missouri 65401, USA

Japanese IQ

SIR — Richard Lynn[1] agrees with me that the value for mean Japanese IQ should be dropped from the 111 of his original article[2] to 106.6, assuming that Americans have made IQ gains at the rate claimed. That rate has now been supported by a massive body of data[3]. He also offers an adjustment for the urban bias of the Japanese sample which, if the missing rural subjects were 7 points below national norms, would reduce Japanese IQ to 104.4.

However, the allowance for urban bias may not be sufficient. The sample not only had no subjects from villages or small towns but also under-represented Japanese from cities less than 50,000 in population[4]. The sample represented only the upper 70 per cent of the population on an urban-rural scale and if the extra missing subjects were even 3 points below national norms, Japanese IQ would fall to 103.9: 106.6(.696) + 100.9(.063) + 96.9(.241) = 103.9.

More important, Lynn makes no allowance for possible bias for socio-economic status (SES) in the sample, a possibility suggested by the fact that no effort was made to stratify for this variable[4]. Mayo Mohs notes that the sample included some elite schools affiliated with private universities[5]. As to the difference an elite SES bias might make, assume that we divided the urban subpopulation of Japan into upper 50 per cent and lower 50 per cent on an SES scale and that 67 per cent of the sample represented the former and 33 per cent the latter. Given a correlation between parent's SES and child's IQ of 0.50 (ref.4), this bias would come to about 2 IQ points: (6 + 6 - 6)/3 = 2. The true mean IQ of Japanese children against current white American norms would drop to 102, at which point the absence of Japanese scores for the verbal component of IQ tests assumes greater significance.

As Lynn points out, evidence on urban-rural differences is mixed[1], and the sample may have been substandard rather than elite for SES. But we just do not know and should await better evidence before giving estimates of Japanese IQ. It is inevitable that textbooks over the next decade will refer to a mean of 111 but the profession should do all in its power to provide an antidote.

JAMES R. FLYNN
Department of Political Studies, University of Otago, New Zealand

1. Lynn, R. *Nature* **306**, 292 (1983).
2. Lynn, R. *Nature* **297**, 222 (1982).
3. Flynn, J.R. *Psychol. Bull.* **95**, 29-51 (1984).
4. Stevenson, H.W. & Azuma, H. *Nature* **306**, 291 (1983).
5. Mohs, M. *Discover*, 20 September (1982).

● Correspondence on this matter is now closed — Editor.

Why abstract?

SIR — In drawing attention to the increasing extent to which abstracts of papers delivered verbally at conferences are being published in scientific journals, M.A. Bray (*Nature* **307**, 206; 1984), bemoaned the fact that large numbers of citations of individual conference abstracts tend to appear among references retrieved by computerized literature searches. You suggest that database compilers might admit as search criteria Boolean specifications such as "not abstract". Would it not be better, however, to consider whether abstract-type papers should be included in such databases in the first place?

Those who favour their inclusion would argue (1) that an abstract from a conference may be the only record of a piece of research, (2) that it gives notice of research results long before they are written up in full, or (3) that since abstracts are often cited in research papers they must be accorded some scientific value.

Those who think abstracts should not be included would argue (1) that they are generally only interim summaries of ongoing research which will be written up in full in due course, (2) that they are often not backed up by fuller reports and so represent only a minimal form of publication or (3) that database compilers should concentrate on conventional primary literature.

The Commonwealth Agricultural Bureaux (CAB), which publishes a large series of abstract journals covering the literature on agriculture and related fields (the consolidated machine-readable database is known as *CAB Abstracts*), is reviewing its policy on conference abstracts. While full papers in conference proceedings will continue to be dealt with like any paper in a primary journal, we need to consider whether individual conference abstracts should be included (the full proceedings will still be cited). Before a final decision is made on this matter, comments from both researchers and database users would be helpful.

J.R. METCALFE
Commonwealth Agricultural Bureaux, Farnham House, Farnham Royal, Slough SL2 3BN, UK

Last amateur

SIR — It is unfortunate that your "100 Years ago" on William Barlow's first paper on the relations of crystal structure to symmetry[1] failed to give him credit by name. Barlow, one of the last in the English tradition of talented self-taught "amateurs", went on to devise a completely original formulation of 230 crystal space groups[2] that was pre-empted by the slightly earlier publications of E.S. Fedorov and Artur Schönflies. His theoretical models of crystal structure, like those you reprinted, inspired W.H. and W.L. Braggs' actual solution of crystal structures by X-ray diffraction. But his valence theory of 1906-10 (with W.H. Pope of Cambridge) was far off the mark, and the new theories of crystal structure passed him by. He did not publish anything further after 1915 although he lived until 1934[3], but he deserves to be remembered as one of the completely original minds of his time.

WILLIAM T. HOLSER
Department of Geology, University of Oregon, Eugene, Oregon 97403, USA

1. *Nature* **306**, 43 (1983).
2. Barlow, W. *Z. Kristallogr.* **23**, 1-63 (1894).

Count down

SIR — In the jargon of radiation detector technology, "count" is used correctly as a verb, but also perversely to describe the event or pulse so counted and, with slightly more justification, for the act of counting for a chosen period and for the result. Thus instructions to a student might read:

"take thirty 30-second counts, record the number of counts for each count, and calculate the average number of counts per second".

That students can nevertheless perform the experiment correctly is an indication that they have been exposed to such misusage from an early age and accept it as normal.

One should count pulses, and restrict the counting of counts to statistical studies of European nobility.

R. F. ENTWISTLE
Department of Physics, University of Otago, Box 56, Dunedin, New Zealand

© 1984 Nature Publishing Group

Origin of Solar System redefined

Professor Thomas Gold has stood the conventional view of the origin of the Solar System on its head and seems certain to provoke astrophysicists to a wave of recalculation, even observation.

NOT for the first time, Professor Thomas Gold from Cornell University has provided astrophysicists with a new agenda, this time on the old question of the mechanism by which the Sun and stars like it were formed. Speaking at last week's meeting on rotation in the Solar System, organized in London by the Royal Society, Gold invited his audience to abandon the conventional framework that the Solar System was formed by rapid condensation within a molecular cloud, perhaps in some tens of thousands of years, and that the Sun and planets were then formed by differential condensation within such a compact mass — and to follow instead his own conviction that the condensation was much slower, perhaps extending over several hundreds of millions of years.

Gold's starting point is the now familiar observation that the angular momentum of the Sun is a small part of the angular momentum of the Solar System as a whole — 1 part in 180 to be precise. The standard explanation is that during the final stage in the evolution of the Solar System when the planets differentiated from the Sun, angular momentum was transferred to the outer objects from the centre, the rotation of which would otherwise be too rapid for stability. Gold's argument is that the Solar System did not form by the initial collapse and subsequent dispersion of angular momentum (and some material), but that slow accretion onto a central object was dominant from the start.

The chief objective of Gold's argument is to explain why the Sun now embodies such a small proportion of the angular momentum of the Solar System as a whole. On his view but also on the conventional view, the total should of course be that of the portion of the molecular cloud from which the whole Solar System is derived less the amount of angular momentum lost during or after its formation. On both views, the original cloud was larger than the present dimensions of the Solar System, so that its rotational speed would have been less than that of the planets (measured in years) or of the Sun (less than a month).

Gold pointed out last week that if Uranus and Neptune, now largely made of carbon, oxygen and nitrogen, began life as much larger objects with a chemical composition more like that of Jupiter and Saturn, the subsequent loss of large amounts of hydrogen and helium would have entailed a substantial loss of angular momentum, in which case the small fraction of the total now embodied in the Sun would have been even smaller in the past.

What can account for this state of affairs? Gold's explanation is simple but ingenious. Let there be some portion of a molecular cloud which, as a consequence of some density fluctuation, is dense enough to be self-gravitating and thus to have a centre of gravity. If the material is also dense enough for intermolecular collisions to be significant, it will begin to accumulate near the centre until, in the course of time, the centre is occupied by an identifiable mass, a proto-star, by which time at least the inner portion of the cloud will have become disk-like. In this condition, the central object will continue to accrete material from the disk, but with an angular momentum appropriate to the orbital speed at its equator. Plainly such a process of accretion cannot avoid the difficulty that a star grown in this way would have a very large angular momentum unless the growing object were more compact than the Sun. But this, Gold said last week, could be the case if the slowly accreting object were statistically degenerate, perhaps even a neutron star. Then, although the equatorial orbital speed will be very large, the increment of angular momentum will be only small. And the central object will come to resemble a main-sequence star only when accretion makes the central temperature and density of the central object great enough for thermonuclear energy production to begin, whereupon the object will quickly grow in size (oscillating in the process) and, like an ice-skater who stretches out his arms, will then spin more slowly.

But what evidence can there be to guide the choice between Gold's accretion scheme and the conventional view? Each entails the production of the central star from a rotating disk of material. Gold last week drew attention to two supporting lines of evidence — the observation that most non-binary stars are also, like the Sun, slowly rotating objects and that the vectors of the angular momentum of the Sun and of its planetary system point in directions which are more than 7 degrees apart. Gold's argument is that this misalignment is simply not explicable if the planets are the means by which a rotating central object sheds excess angular momentum, but that slow accretion from what might well be a puckered disk can easily accommodate the discrepancy.

It is too soon to know where all this will lead. The ideal would of course be that the same model of slow accretion should also account for the formation of binary stars, which may presumably be the products of the fission of degenerate stars rotating too quickly for stability, but that arithmetic has not yet been done. What is clear is that for single slowly accreting stars, the rate of heat conduction through the degenerate material of the growing but inert proto-star will limit the rate of accretion that can be accommodated without igniting thermonuclear fusion, whence the guess that the formation of visible stars will be slow. An obvious difficulty for the new model, raised last week, is the homogeneity of age of stars in globular clusters and that it may allow too many black holes to form.

Gold himself last week acknowledged that the expansion of a degenerate star would probably not be sufficient to account for the slowness of rotation of the Sun, perhaps by a factor of five, although he argued that mechanisms involving the outward transfer of angular momentum by the agency of magnetic fields (invoked on the conventional view) are more than adequate to bridge this gap. One intriguing possibility is that it may be possible to identify proto-stars in this mould by looking in molecular clouds for Doppler-shifted emission from material travelling with the equatorial orbital speed just above the surface of a rotating degenerate star.

The planetary implications of Gold's model are also potentially fruitful. One possibility is that late in the accretion of the central star, the residue of the molecular cloud might consist of a sequence of rings not all in the same plane; Gold argues that such a system could account not only for the formation of Oort's cometary belt and for the occurrence of chemically inhomogeneous meteoritic material but also for the suggestions that the material of which the Solar System is made has been contaminated with material derived from at least two supernova explosions whose products are not distributed evenly.

Last week, Gold himself was obviously shy of carrying such arguments too far, if only because that would accommodate an *ad planetum* explanation of each object in the Solar System. Diffidence on that account, natural enough, is not however an argument against Gold's model, which is certain in the next few months (or weeks?) to have people in the field rushing to their computer programs. **John Maddox**

© 1984 Nature Publishing Group

Astronomy

An infrared view of the Universe

from Joseph Silk

IT is said that the holy grail of infrared astronomy is the discovery of a star being born. The Infrared Astronomical Satellite (IRAS) has brought astronomers closer than ever before towards attaining it.

The satellite experiment is the result of a collaboration between British, Dutch and US astronomers, and was intended to provide a systematic, unbiased and sensitive survey of the entire sky over four infrared bands, at 12, 25, 60 and 100 μm. The advantages of conducting an infrared survey above the Earth's atmosphere are twofold: first, only in space can a telescope be sufficiently cold to overcome the thermal pollution that fogs the infrared sky as seen from the ground; and second, it makes it possible to peer more deeply into the Universe. Although the satellite, launched on 23 January 1983, exhausted its cryogenic helium supply barely 10 months later, thereby terminating the survey, the accumulated data will occupy astronomers for years to come. The first scientific results appear in the 15 March issue of *Astrophysical Journal Letters;* here I discuss some of the highlights.

Preliminary results of a minisurvey over some 900 square degrees reveal nearly 9,000 infrared sources, according to Rowan-Robinson *et al.* Away from the plane of our Milky Way galaxy, more than half of the sources at 12 and 25 μm can be identified with bright stars. About a quarter of the sources seen at longer wavelengths are identified with known galaxies while many others are more local objects embedded in dust clouds — perhaps newly forming stars. Within 10° of the galactic plane, there is so much activity associated with star formation that it becomes exceedingly difficult to pick out individual objects from the survey.

Study of the sources that can be identified with known objects has proved rewarding. A characteristic feature of many regions of star formation is the presence of energetic outflows, which can be detected both directly as high-velocity molecular gas and by the presence of shock-excited gas clumps — known as Herbig-Haro objects — which have been conjectured to be the remnants of the interaction of a high-velocity wind with a dense inhomogeneous molecular cloud. Through infrared observation it is possible to probe the nature of the heavily embedded objects that drive the outflows.

Emerson *et al.* studied two candidate driving sources in the dark cloud L1551, which is associated with three Herbig-Haro objects and has a well-mapped bipolar flow. One source (IRS-5) had previously been identified with a low-mass protostar that drives the molecular gas outflow. The other is 0.1 pc away from IRS-5 and appears to be a dust-embedded precursor of a low-luminosity pre-main-sequence star. Star formation is occurring at more than one place in the L1551 cloud, as predicted if the parent cloud has undergone fragmentation. The situation resembles that seen in sites of massive star formation: low-mass stars may form by a similar mechanism. Another pair of Herbig-Haro objects, HH46 and HH47, appear to have been ejected from a nearby dark globule, which contains a similar dust-embedded precursor of a low-mass star.

That formation of low-mass stars is endemic to dark globules has been confirmed in more detailed studies of dark clouds. Two compact sources of radiation were discovered within the dense core of the cloud Barnard 5. One is a newly formed protostar of solar mass; the second is much cooler and may be a density fluctuation in the very earliest stage of collapse. Two other newly formed stars of solar mass, still enshrouded in dust shells, were found nearby. Thus Barnard 5, an apparently quiescent globule of dense molecular gas, has now been observed to contain as many as four protostars. The stars may have formed by the fragmentation of the dense core of Barnard 5 and are now responsible for providing enough energy input to stabilize the cloud against further collapse. The dark cloud complex Chamaeleon I was found to contain many newly formed stars. Some sources have been identified with known pre-main-sequence objects, but a considerable proportion have no optical counterpart: presumably, they are deeply embedded in the cloud. Cool sources were detected only at the longer wavelengths and appear to surround the main cloud. They seem to be associated with small globules and may represent the earlier stages of collapse. Perhaps a hundred thousand years from now, protostars of solar mass will have emerged from cocoons in the globules.

Clearly IRAS has produced the best links, as yet, between molecular clouds and newly formed stars. But it has also allowed a fascinating insight into the nature of galaxies because almost all the energy produced by newly formed stars is emitted from their placental gas and dust shrouds at infrared wavelengths. The infrared properties of galaxies determine the distribution, energetics and overall rate of star formation which is invaluable information for understanding galactic evolution.

In a study of a complete optical sample of 165 nearby galaxies, de Jong *et al.* verify that infrared emission is generally correlated with blue luminosity, which is associated with massive young stars. Predominantly spiral galaxies have been detected by IRAS; normal ellipticals in the sample were not detected.

One surprise is that the range of infrared-to-blue-luminosity varies considerably from galaxy to galaxy. While the Andromeda galaxy, for example, is a very weak infrared source, emitting only 3 per cent of its blue light in the infrared, other more active spirals emit up to five times as much infrared as blue light. The weakly emitting spirals have rather cool dust (at a colour temperature of ∿25K), while the more luminous infrared galaxies are often warmer (∿50K), suggesting an enhanced formation rate of massive O stars surrounded by molecular clouds.

It might be predicted that a sample of galaxies selected from the infrared survey would contain the most infrared-luminous galaxies. Indeed, Soifer *et al.* identify nearly 100 spirals that have infrared luminosities greater than or comparable to their luminosities in the blue. The wide range in infrared luminosities is probably the result of variation in dust content from galaxy to galaxy, as well as variations in star formation rate. The very luminous infrared galaxies are only a small proportion of all galaxies. However, a surprisingly large fraction of the infrared-selected sample are found to have neighbouring galaxies nearby. Theory has suggested that the tidal interaction of a galaxy with close neighbours can trigger outbursts of star formation, and the IRAS data provide tantalizing evidence for this conjecture. Moreover Young *et al.* have found that many of the infrared sources detected in the Hercules cluster of galaxies are spiral galaxies. The most luminous galaxy has a warped disc and has been catalogued as an interacting galaxy. Its enhanced infrared luminosity is probably a result of a burst of star formation triggered by the interaction: tidal forces would have compressed molecular cloud complexes and initiated collapse.

Since some galaxies can be such spectacular emitters in the far infrared, infrared emission might reveal the presence of galaxies where optical observations are inconclusive. This seems to be the case for quasars. In a comparison of the infrared data on radio-loud and radio-quiet quasars, Neugebauer *et al.* find the infrared continua of the radio-loud quasars are consistent with the spectrum expected from synchrotron emission by relativistic electrons, as extrapolated from the radio-frequency region. By contrast, the two radio-quiet quasars observed reveal an excess over the power-law component at 100 μm, suggesting a contribution from an underlying spiral galaxy that is vigorously forming stars and is bright at 100 μm. Indeed there is evidence at 1–2 μm that the infrared emission around both radio-quiet quasars is extended, consistent with the presence of a surrounding galaxy. The infrared 100 μm luminosity for the underlying galaxy is

© 1984 Nature Publishing Group

similar to that of active spiral galaxies.

One of the most spectacular results from IRAS is the discovery by Auman *et al.* of an infrared excess, extending to 20 arc s at 60 μm, around the bright star Vega. The most likely origin of the infrared emission is thermal radiation from solid grains, at least 1 mm in radius — smaller grains would have spiralled into Vega by the Poynting–Robertson effect — and heated by light from Vega to about 85K. The grains must lie in a shell or ring at a distance of about 85 AU from Vega and in orbit around the star. These particles cannot have been present since the time of formation of Vega, some $1–2 \times 10^8$ yr ago, since they are much larger than typical interstellar grains (which are at most 0.5 μm in size). Clearly they have grown considerably since Vega formed. Since our Solar System is about 4.5×10^9 yr old, it is tempting to infer that the shell around Vega represents some intermediate stage in the formation of a planetary system.

Another major surprise from IRAS is the discovery of interstellar cirrus. Low *et al.* report extended patches of far-infrared emission, seen predominantly at 60 μm and 100 μm, both high above the galactic plane and in the ecliptic plane. Three components have been identified: one associated with neutral hydrogen and dust concentrations at very high galactic latitudes; another in the ecliptic, associated with the interplanetary medium and apparently running continuously around the Solar System; and a third — the coldest — which correlates poorly with any known structure in the Galaxy or Solar System. Degree-size patches of 100 μm emission are seen and could represent either dust clouds in the outer Solar System, or a novel component of the interstellar medium. IRAS observations, taken six months apart, should be capable of resolving this enigma.

Finally, the IRAS minisurvey has revealed some very intriguing unidentified objects that may be of a fundamentally new type. Houck *et al.* identify nine point sources that are bright at 60 μm but have no obvious identifiable optical counterpart brighter than 18.5 mag. The intrinsic infrared luminosity exceeds that in the blue band by at least an order of magnitude. Of course, these sources may not be extragalactic or even outside the Solar System. Nevertheless, their discovery is one of the potentially most important accomplishments of IRAS. Radio surveys led to the discovery of quasars and pulsars; X-ray surveys came up with X-ray binaries and bursters; now the race will begin to explain the unidentified sources in the first all-sky infrared survey.

During the next decade both the European Space Agency and the National Aeronautics and Space Administration have plans for a sequel to IRAS. Both the European (ISO for Infrared Space Observatory) and the American (SIRTF for Shuttle Infrared Telescope Facility) experiments are designed to be operated as true observatories for infrared astronomy, carrying a variety of focal-plane instruments. Cryogenically cooled 1-m class space telescopes will be able to study the infrared sources discovered by IRAS at more than 100 times the sensitivity and spectral resolution and provide detailed pictures of even the faintest IRAS sources. This quantum leap in infrared capability will do for the infrared astronomers what the Einstein Observatory did for X-ray astronomers. Only then will we see more than the tantalizing glimpse of the infrared sky that IRAS has already provided. □

Joseph Silk is at the Astronomy Department, University of California, Berkeley, California 94720.

Radiography of the living brain

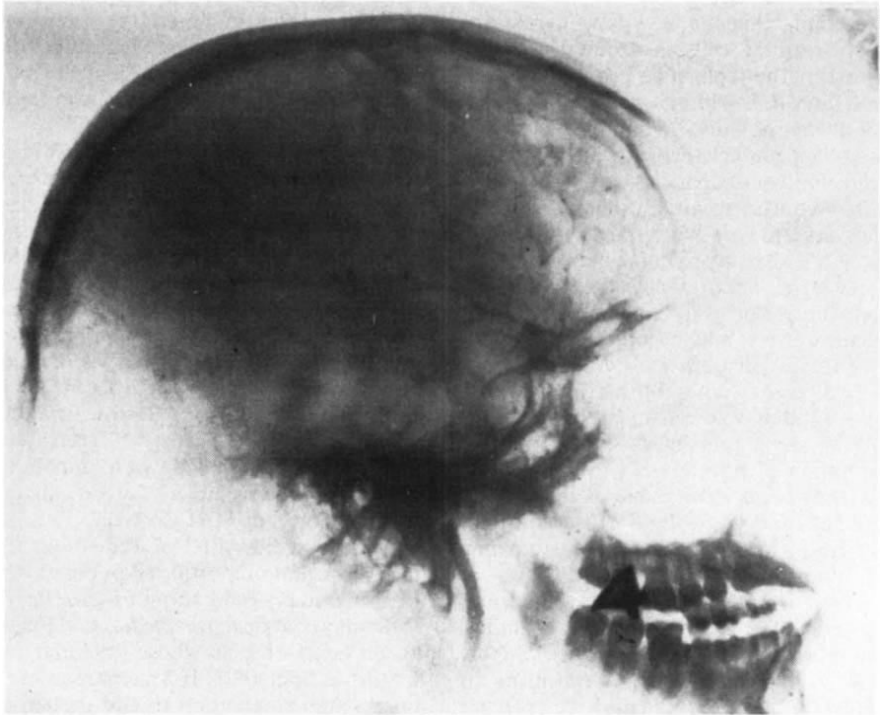
It was not until thirty years after the first X-ray images of the skull that the first cerebral angiograms, obtained by Egas Moniz and shown here, revealed the distribution of tissue, blood and water in the brain.

A neurologist at Santa Maria Hospital in Lisbon, Moniz planned to inject a contrast agent in order "to make the diagnosis of the localization of the majority of [brain] tumours through the alteration of the normal arterial pattern of the cerebrum". The idea was not new: as early as 1896 E. Haschek and O.T. Lindenthal in Vienna studied the hand of a corpse through injection of Teichmann's mixture, a toxic amalgam of petroleum, quicklime and mercuric sulphide. But to perform angiography on a living patient, Moniz had to find a less hazardous contrast agent and a way to inject it so that good photographs could be taken before it was heavily diluted by the blood flow.

The method he adopted consisted of exposing the internal carotid artery, injecting the contrast agent, immediately placing a ligature on the artery and then releasing it a split second before taking the radiogram. With his third subject Moniz achieved success at an exposure of 0.25 s using a 25 per cent solution of sodium chloride. "At that unforgettable hour, the afternoon of 28th of June 1927," wrote Moniz later, "all our attention was concentrated upon the first arteriogram. We remember with pleasure the work done, . . . the culmination of an obsession to realize a long-standing project. On the film [shown below] the cerebral vessels were seen, but deformed through the presence of the tumour. The internal carotid artery was pulled forward. . . The sylvian artery was displaced upward from its normal position."

Within the next decade another thousand diagnoses by means of angiography had been successfully conducted and the technique had been applied by Reynaldo dos Santos, Moniz and others to obtain arteriograms of other parts of the body. Ironically, Moniz was to share the 1949 Nobel Prize for Medicine on the strength of originating the technique of prefrontal lobotomy, a practice later discredited, whereas his lasting contribution to cerebral angiography went unrecognized at the time. **Jon Darius**

Selected from Beyond Vision *by Jon Darius of the Science Museum, London (to be published in April by Oxford University Press). Radiogram reproduced by courtesy of Antonio Coito of the Santa Maria Hospital in Lisbon, Portugal.*



© 1984 Nature Publishing Group

Cell proliferation

Paths to immortality and back

from T.B.L. Kirkwood

RECENT exciting developments in the characterization of cellular and viral oncogenes show that the immortalization of mammalian cells is an important feature of their progression from normality to malignancy[1,2]. This finding makes natural sense; populations of normal cells exhibit only finite division potential when grown in culture, and it is commonly held that there is an intrinsic limit to cell proliferation that must be overcome *in vivo* before a cancer can be formed. (But immortality is not absolutely required; if the potential for cell division is great enough a lethal tumour can be found without it.) At the same time, it is widely believed that the limited growth of cells such as fibroblasts is a contributing factor in the process of ageing[3].

If both these ideas are true, then cancer and ageing are intimately connected at the cellular level, and if we could but understand the mechanism by which one process occurs, we should be better placed to understand the other. Behind the simple dichotomy between finite and infinite cell population growth, however, lie complex issues.

First, there are two quite different types of theories to explain why cells stop proliferating. One postulates a rigourously programmed series of events specifically aimed at terminating cell division; the other, that cells cease growing because they accumulate random defects, particularly among macromolecules. The two mechanisms call for quite different explanations of how cells can escape the ageing process and become immortalized[4], although testing for this difference is by no means easy.

Second, while the growth behaviour of a population of cells is necessarily determined by the replicative properties of the individual cells within it, the relationship is not always obvious. If immortal cells give rise to mortal cells which undergo only a finite number of divisions before they die, a culture consisting initially of immortal cells may nevertheless have only a finite lifespan[5]. This surprising outcome arises because the immortal cells are 'diluted' by their mortal progeny and, in appropriate circumstances, may become so scarce that they are lost altogether.

Two recent studies, both by groups with long-standing interests in the problems of cellular ageing, suggest that the transformation of normal cells to immortality may not be as irreversible as is generally thought, and throw light on some of the difficulties in interpreting the phenomenon of immortalization.

Pereira-Smith and Smith[6] used fusions between different combinations of mortal and immortal somatic cells to see whether cellular immortality is dominant or recessive. When normal cells were hybridized with three simian virus 40 (SV40)-transformed cell lines and with a variety of cell lines (including HeLa) derived from malignant tumours, in all cases the phenotype of immortality was recessive, although variant immortal cells did arise at low frequency (around 1 in 10^5) within some of the non-proliferating hybrid populations. These results confirm and significantly extend earlier reports of the dominance of mortality over immortality in human cell hybrids[7-9].

More strikingly, Pereira-Smith and Smith have also found that in fusions between some pairs of immortal cell lines, growth is finite but that in others, all hybrids cells can divide indefinitely. This surprising result is interpreted as support for the hypothesis of programmed cessation of cell division, the argument being that immortality can arise from various distinct dysfunctions in the programme and that, among hybrids between cell lines immortalized in different ways, complementation can restore the operation of the programme for mortality. I will, however, explain that this finding does not militate against the alternative hypothesis of random damage as strongly as Pereira-Smith and Smith have claimed; in any case, for a trait as complex as immortality, the classical concepts of dominance and recessiveness may be too narrow.

The second study, by Huschtscha and Holliday[10], details the events surrounding transformation by SV40 of the human diploid cell strain MRC-5, comparing two permanent cell lines obtained therefrom. The lines differ in several respects, including morphology, modal chromosome number, glucose requirement and, in particular, stability of the immortal phenotype.

One of the lines (called MRC-5V1) is unusually unstable in that the growth of some sub-lines slows over a period of many population doublings, the cultures dying out in a manner closely resembling the behaviour of normal cells. By recovering cells frozen in liquid nitrogen, the line was kept growing through 750 population doublings, raising the question whether the immortal phenotype is carried only by a small sub-population of cells which may on occasion have been lost. Regrettably, Huschtscha and Holliday were unable to test this hypothesis directly by growing individual clones of MRC-5V1 cells, but independent studies with HeLa and other cell lines show that immortal cell populations, even after many generations of selection *in vitro*, may contain a sizeable fraction (up to 40 per cent) of cells whose potential for division is limited[6,11]. If a similar, or even more extreme, situation should pertain *in vivo*, which seems quite plausible because selection for immortality *per se* will not begin until normal clonal lifespan is exhausted, it could explain why it is not always easy to establish permanent cell lines from tumour biopsies and why tumorigenesis itself sometimes seems a hit-or-miss phenomenon.

So what of the mechanism limiting normal cell division? If it contributes to the ageing of the organism as a whole, it must be understood in an evolutionary context as much as in any other[4]. The popular view of ageing as an adaptive mechanism to prevent overcrowding and promote species' adaptability, which lends direct support to the hypothesis of programmed cell death, is regarded by most evolutionary biologists as untenable[12]. On the other hand, there is quite good support for the view that ageing is the consequence of the energy-saving strategy of not repairing cellular damage too well[4,13]. This 'disposable soma' hypothesis suggests that somatic cells are switched to a repair level which is optimally efficient from the organism's point of view even if less than optimal for the cell. If this is the case, transformation of a cell to immortality can be understood as a product of clonal selection acting on rare mutations or epigenetic changes which enhance the capacity of a cell to cope with damage, either by repairing it more effectively or by diluting it with more rapid cell division. It is possible that such a change might be 'recessive' and, therefore, consistent with the data of Pereira-Smith and Smith[6].

The advances made in recent months in understanding cell immortalization suggest that a new path to unravelling the mechanisms of carcinogenesis and cell ageing may have been opened up. In working our way down this path, it will be helpful to keep the competing hypotheses of both these fields firmly in mind. Looking back nearly a hundred years, it is amusing to note the sharp prejudice of Vines[14], who wrote of Weismann's theory of cell ageing that it is "absurd to say that an immortal substance can be converted into a mortal substance". Poor Professor Vines has been proved wrong. The conversion appears to work both ways, and with any luck we may soon know why. □

T.B.L. Kirkwood is at the National Institute for Medical Research, Mill Hill, London NW7 1AA.

1. Cairns, J. & Logan, J. *Nature* **304**, 582 (1983).
2. Land, H., Parada, L.F. & Weinberg, R.A. *Science* **222**, 771 (1983).
3. Hayflick, L. *Scient. Am.* **242**, 42 (1980).
4. Kirkwood, T.B.L. & Cremer, T. *Hum. Genet.* **60**, 101 (1982).
5. Kirkwood, T.B.L. & Holliday, R. *J. theor. Biol.* **53**, 481 (1975).
6. Pereira-Smith, O.M. & Smith, J. R. *Science* **221**, 964 (1983).
7. Bunn, C.L. & Tarrant, G.M. *Expl Cell Res.* **127**, 385 (1980).
8. Muggleton-Harris, A.L. & De Simone, D.W. *Somat. Cell Genet.* **6**, 689 (1980).
9. Pereira-Smith, O.M. & Smith, J.R. *Somat. Cell Genet.* **7**, 411 (1981).
10. Huschtscha, L.T. & Holliday, R. *J. Cell Sci.* **63**, 77 (1983).
11. Martinez, A.O. *et al. In Vitro* **14**, 996 (1978).
12. Medawar, P.B. in *The Uniqueness of the Individual* 2nd edn, 28 (Dover, New York, 1981).
13. Kirkwood, T.B.L. *Nature* **270**, 301 (1977).
14. Vines, S.H. *Nature* **40**, 621 (1889).

© 1984 Nature Publishing Group

Laser chemistry

Chaos in multiple photon excitation of molecules

from Peter Knight

THE discovery in the early 1970s by Isenor and Richardson (*Appl. Phys. Lett.* **18**, 224; 1971) and others that some polyatomic molecules could be dissociated easily by multiple infrared photon absorption seemed to offer the exciting prospect of using cheap infrared gas laser photons to catalyse or steer unimolecular chemical reactions. But by the early 1980s it was clear that most polyatomic molecules seemed not to be coherently excited by infrared lasers. Instead all that mattered, it seemed, was how much energy the pulsed laser could dump into the molecule, which rapidly thermalized the excitation. Such absorption is fluence dependent instead of intensity dependent. An industry has grown up among photochemists in which this thermalization is discussed in terms of statistical kinetic theory. In this theory, molecular excitation takes place in a ladder of energy levels of a fundamental anharmonic pump mode embedded in a much larger number of background 'quasi-continuum' modes (Fig. 1). Intramolecular coupling acts to redistribute, or thermalize, the laser energy into these modes. Now Ackerhalt, Galbraith and Milonni (*Phys. Rev. Lett.* **51**, 1259; 1983) have invoked the fashionable concept of chaos to suggest an entirely new reason for the fluence dependence and apparent statistical behaviour of most polyatomic molecules excited by infrared laser.

Using realistic models of multiple photon excitation, Ackerhalt *et al.* have demonstrated that linear absorption with 'noise' superimposed can result from the intensely nonlinear excitation of a polyatomic molecule. They do not need to invoke statistical or thermal redistribution mechanisms. Their analysis seems general and provides not only an explanation for their specific problem in laser chemistry but also a concrete example of the occurrence of chaos in a microscopic system.

The model employed by Ackerhalt, Galbraith and Milonni retains the description of the infrared vibrational mode as an anharmonic excitation pumped by the laser field. The anharmonic mode is coupled to a set of many background modes. These are traditionally treated as dense enough to form an irreversible sink into which excitation leaks. Ackerhalt *et al.* instead treat them as uniformly distributed harmonic modes rather as in the theory of radiationless transitions. The anharmonic mode coupling to the background is then reversible. Energy is transferred into the various background states so the molecular excitation dephases but eventually rephases and couples back to the pump

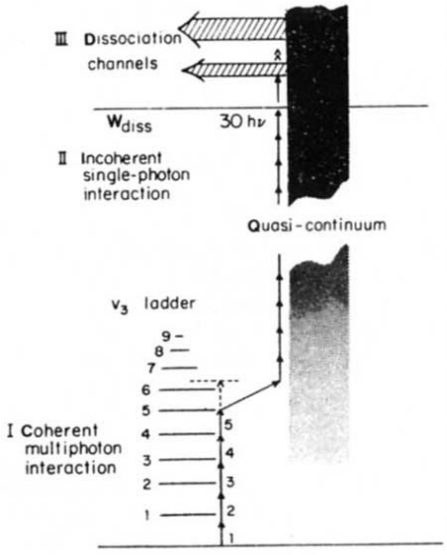


Fig. 1 Pump mode levels and background quasicontinuum of polyatomic molecule (from Black *et al. Phys. Rev. Lett.* **38**, 1131; 1977).

mode. In this way the level of excitation in the molecule is periodically kicked. As a result, the evolution of the molecular excitation rapidly becomes chaotic.

Figure 2 shows the number of photons absorbed in their model as a function of time. The boxes show the steady-state absorption in each interval between kicks. The dotted curve is the dynamics predicted from the full analysis including the kicks. The molecular and laser parameters were chosen to be typical of a real molecule. The pumped mode may reach steady state in each interval between kicks and stop absorbing, but the background modes continue to take up laser energy using the pumped mode as an intermediary. The result is linear absorption with noise superimposed. The chaos appears to average over the induced coherences to produce the linear rate observed. That the evolution is truly chaotic (in the sense of exponential sensitivity to initial conditions) was checked by calculating the Fourier transform spectrum and the maximal Lyapunov exponent. Chaotic motion leads to a broad-band spectrum whereas periodic motion has a discrete spectrum.

The model has been elaborated to investigate whether the chaotic evolution is generic rather than specific to the non-rotating anharmonic model. Intramolecular relaxation in the background modes and the removal of the assumption of equally spaced background modes still allows chaotic evolution over a wide range of driving parameters. Recent work by Galbraith, Ackerhalt and Milonni (*J. chem. Phys.* **79**, 5345; 1983) has considered a model with no background modes at all but with a vibration–rotation coupling which generates nonlinear equations of motion. Again chaos is observed.

Many recent observations of chaos occur in artificial feedback-driven models or models driven outside of their region of validity. Ackerhalt, Galbraith and Milonni believe multiple photon absorption leads to chaotic excitation of polyatomic molecules as a generic feature for realistic values of molecular parameters. The signature of chaos, they claim, is a fundamental aspect of the physical process. It is ironic that coherent multiple photon absorption, important only for intense laser fields, could be prevented by chaotic dynamics created by strongly driven anharmonicities induced by that same field. □

Peter Knight is a Lecturer in Physics in the Blackett Laboratory, Imperial College, London SW7 2BZ.

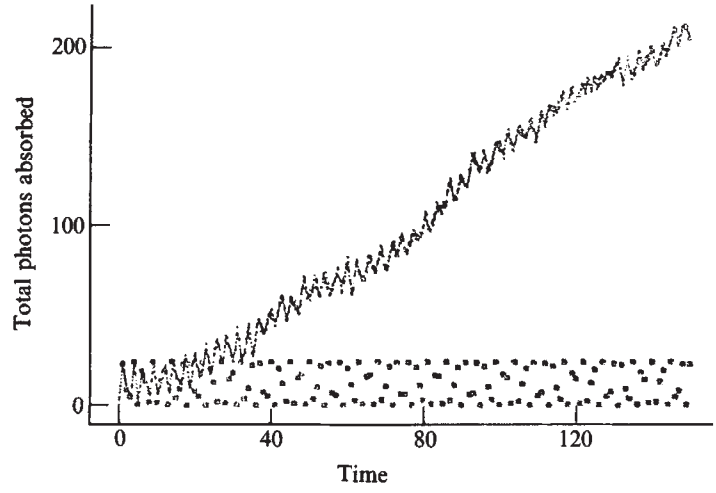


Fig. 2 Total photons absorbed by polyatomic molecule in new chaotic model. Boxes indicate steady states between rephasing kicks (Galbraith *et al. J. chem. Phys.* **79**, 5345; 1983).

© 1984 Nature Publishing Group

Vision

Visual hyperacuity

from Oliver Braddick

A RECENT symposium* at Cambridge set out to review advances in our understanding of 'visual hyperacuities' — spatial judgements which imply a positional accuracy in vision a good deal finer than the spacing of the mosaic of photoreceptors in the retina.

The term 'hyperacuity' was originally coined by G. Westheimer (University of California, Berkeley), who described at the symposium the wide variety of visual tasks that demonstrate the use of fine positional information. It is now generally understood that there is no paradox about positional judgements that are finer than the receptor mosaic: the patterns are imaged over many receptors and, by virtue of neural interactions, their effects spread to many parallel channels in the visual system. Thus there are many opportunities for integration of the available information, but the question remains: what aspects of this complex pattern of information are available to, and used by, the decision processes that underly our visual judgements?

The judgement of equal spacing in a set of parallel lines has been the focus of several recent studies, partly because the pattern can be analysed in terms of its effect on spatial frequency bandpass mechanisms, which are believed to play an important part in early visual processing. S. Klein (University of Houston) outlined how deviations from equal spacing produce characteristic patterns in an array of mechanisms, each responding to a limited region of space and spatial frequency — dubbed 'viewprints' by analogy with acoustic 'voiceprints' used to represent variations in time and sound frequency. B. Madden (University of Rochester) developed a theory of this kind in more detail: his four-line display has a spatial frequency spectrum with certain zeroes, and a channel tuned to one of these will detect contrast when a threshold deviation from equal spacing is introduced. The thresholds found are consistent with those for contrast detection at the spatial frequency concerned and, as the theory would predict, superimposing a contrast grating of that frequency distorts the hyperacuity judgement. An implication of this approach is that hyperacuity judgements do not necessarily require the decision processes to have available explicit positional information.

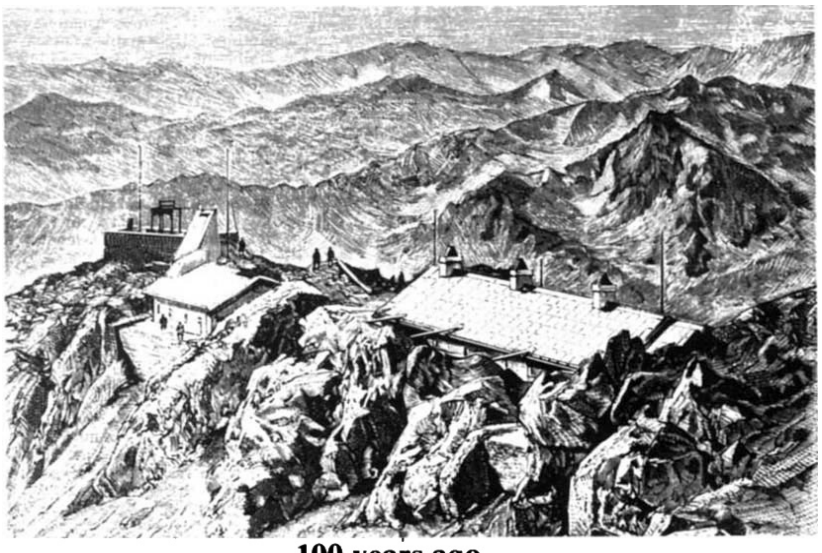
If, however, fine positional information is explicitly represented in the nervous system, what is the nature of the representation? Barlow has suggested that the image sampled by the photoreceptors might be recreated by a literal interpolation, perhaps performed by the cells receiving visual input in layer IVc of the striate cortex. Indeed, individual cells do provide a statistically reliable signal for pattern differences in the hyperacuity range, but the form of their receptive fields does not seem particularly well adapted for providing an interpolated fine-grain spatial representation (A. Parker and M. Hawken, University of Oxford).

Some form of interpolation is also suggested by demonstrations that hyperacuity judgements are possible even for images that are presented as a set of clearly resolvable samples. M.J. Morgan (University College, London) pointed out that no explicit interpolation is necessary if the image is separately processed by low-spatial-frequency channels that are insensitive to the sampling frequency, and high-frequency channels that signal the discreteness of the samples. Why, then, should the 'interpolation' of a sampled image ever fail? Morgan's experiments show that high spatial frequencies introduced by sampling can interfere with the low frequencies required to make the vernier judgement, in a way that makes it difficult to maintain that human observers can separately inspect their different spatial frequency channels. These effects seem analogous to the well known difficulty of recognizing block quantized images (the 'Abraham Lincoln' effect). Morgan and R.J. Watt propose that hyperacuity judgements (and pattern perception more generally) are based on 'primitives' derived from a non-linear but analogue combination of channel outputs.

The detection of visual displacement and of stereoscopic disparity also shows hyperacute sensitivity to position. It seems unlikely, however, that the detailed mechanisms suggested for pattern hyperacuities can be extended to motion and stereo. Thus stereo performance seems to depend on integrating information along the length of the disparate contours (S. McKee and C.W. Tyler, Smith Kettlewell Institute, San Francisco), which is relatively unimportant in vernier acuity. P. Heard and R.L. Gregory (University of Bristol) generated apparent displacements of contours by manipulations of their luminance profiles and found that the apparent extent of movement followed quite a distinct function from the contour location determined in vernier and stereo tasks.

The visual system, then, appears to have a limiting sensitivity to position or configuration of a few seconds of arc, but the mchanisms which use this precision in the visual information are quite different in detail for different tasks. Ultimately, one of the factors determining this precision must be the accuracy with which the system knows the location of the photo-receptors. J. Yellot (University of California, Irvine) showed that the photoreceptor mosaic in the central fovea has a striking regularity. The potential penalty for this regularity is aliasing effects, but the high spatial frequencies which would produce them are normally removed by the optical limitations of the eye. In more peripheral regions of the retina, the mosaic is coarse enough that aliasing could produce visible degradation, were it not that the photoreceptor array has just enough irregularity to avoid it. Differences between peripheral and foveal vision were also considered (M. Fahle, University of Tubingen, and C. Stephenson, University of Cambridge): structural differences between these areas may well provide important clues to the mechanisms under-

*The symposium on 'Biological and engineering aspects of visual hyperacuity and depth perception', sponsored by the Rank Prize Funds' Optoelectronics Committee, was held on 9-11 January 1984



100 years ago

General view of the Pic du Midi Observatory. From *Nature* **29**, 432, 6 March 1884.

© 1984 Nature Publishing Group

lying hyperacuity and other visual performances.

At the meeting, it became clear that modern display technology, particularly in the instrumentation and simulation of advanced aircraft, is aiming at remarkably high standards of realism. P. Waddell (University of Strathclyde) described the long history of attempts to make three-dimensional television and radar. The attempts have often started from an assumption that binocular disparity is a necessary cue for realistic depth, but as demonstrated by K. Boff (US Air Force Aerospace Medical Research Laboratory, Ohio) and M. Roberts (Rediffusion Simulation Ltd), there is an abundance of cues to depth perception, particularly when movement is involved. I. Shanks (Unilever Research) suggested that part of the solution to producing greater depth realism might lie in the removal of surface cues that interfere with perceived depth, and an optical system was demonstrated that enhanced depth in a television display by this method.

Psychophysical investigations can help the display designer by measuring the sensitivity of vision to changes in image quality (G. Burton, Royal Armament Research and Development Establishment, Sevenoaks). The hyperacuities provide a challenge to the engineer in the stringent demands they place on display quality, but also an opportunity in the precision of the information they can potentially convey to the viewer. We may hope that advances in our knowledge of vision will make these opportunities possible to grasp. □

Oliver Braddick is in the Department of Experimental Psychology, University of Cambridge, Cambridge CB2 3EB.

Archaeology

How to spot a fake azilian pebble

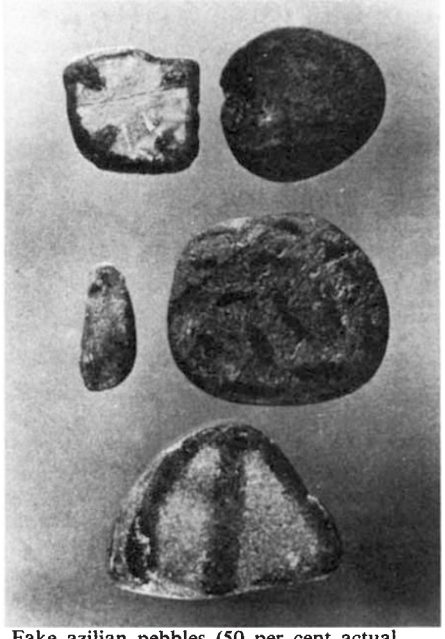
from Paul G. Bahn

A COMPLETE analytical study of every available azilian pebble from collections in Western Europe has helped clear up the long dispute about the authenticity and significance of these flat decorated pebbles found in the Azilian facies of the Final Palaeolithic. His study leaves Claude Couraud, a researcher at the Musée des Antiquités Nationales in Paris, in no doubt that museum collections contain a number of fake pebbles and that the decorations, painted in red ochre, are not just random[1].

The Azilian facies was first discovered by Piette in the Pyrenees in the 1880s. It was thought to be characterized by small flat perforated harpoons and by small flat pebbles decorated with designs in red ochre. Although a few decorated pebbles had been found earlier in other sites, it was the hundreds encountered by Piette in the cave of Mas d'Azil (Ariège, France) which served to identify the 'culture' and which started a long dispute as to their authenticity[2–4].

Decorated pebbles have been found in 28 sites in France, 5 in Spain, 3 in Italy and 1 in Switzerland. However, of the almost 2,000 known over 1,400 are from the Mas d'Azil[5]. The exact total from this site can never be known because there have been losses; the finds are scattered in at least 16 museums and private collections; and there are many fakes.

Although most of the pebbles found in Piette's excavations are undoubtedly genuine, his workers inevitably sometimes took advantage of the system whereby they were paid for their finds. It was some years later that a real trade in fake pebbles was established[2] and, surprisingly, one of its principal victims was the abbé Henri Breuil, doyen of French prehistorians: he bought quite a number of pebbles, and



Fake azilian pebbles (50 per cent actual size) from the Mas d'Azil, Ariège, France. Bégouën collection. Photograph by R. Bégouën.

even published some which are now known to be forgeries[6]. In his old age, there was a final unfortunate episode when he published an 'azilian pebble', found on a Mediterranean beach[7], which was, in fact, a sea-worn fragment of modern mosaic.

Genuine azilian pebbles seem to have been carefully selected for shape and size: most are oval or oblong, up to 8 cm long, up to 3 cm wide and up to 1 cm thick; schist is the most common material. Scholarly suggestions of their function have ranged from 'art for art's sake'[3] to totemic 'churingas'[8] or markers for a game[9]. Piette himself imagined that the marks were numbers, symbols, pictographs and even an alphabet, so that the Mas d'Azil cave represented a huge school[10].

One of Couraud's most important conclusions is that particular combinations of signs recur frequently, and that certain numbers of dots are constantly repeated. For example, there are 16 simple sign-types (mostly dots and transverse or longitudinal lines), but only 41 of the many possible binary combinations have ever been found. Some signs are found in binary associations, while others are mutually exclusive. Such non-random decoration suggests that some sort of 'syntax' is at work. As for repeated numbers, there is a predominance of 21 and 29 or their multiples; and, in the wake of Marshack[11] and Thévenin[12], Couraud has speculated that some form of cyclic notation is involved, perhaps lunar in nature[13].

In addition, Couraud establishes criteria that distinguish authentic pebbles from fakes: where there is calcite on the paint, the authenticity is not in doubt. Fakes display unnatural shades of red, 'newness' of paint and/or pebble, traces of recent embellishment (such as crayon marks), paint inside fractures and clumsy or otherwise aberrant painting. By these criteria he has been able to expose a number of fakes in many collections (see plate), including that of the Musee des Antiquités Nationales.

Most recently, Couraud has examined the 13 pebbles at the British Museum (10 of which were bought from Breuil in 1929) and 12 at the Cambridge University Museum of Archaeology and Ethnology. According to his criteria, 3 of the Cambridge pebbles are fakes; and there are 10 fakes at the British Museum, including 9 of the Breuil collection[14]. It is unfortunate that a postcard on sale at the British Museum features 7 azilian pebbles, 5 of which are fake according to Couraud's analysis.

It has not yet been possible to study azilian pebbles that have found their way into museums and collections outside Western Europe, although it is known, for example, that there are some in the USA. Any information concerning their number and whereabouts would be most welcome. □

Paul G. Bahn can be contacted through the School of Archaeology and Oriental Studies, University of Liverpool, Liverpool L69 3BX.

1. Couraud, C. *L'Art Azilien: Origine, Survivance, Comparaisons* (Thèse de 3e Cycle de l'EHESS, Paris, 1980) (*Gallia Préhist.*, in the press).
2. Bégouën, H. *Bull. Soc. prehist. fr.* **26**, 392 (1929).
3. Péquart, M. *12th Congr. Préhist. fr.* 547 (1936).
4. Bonavia, D. *An. Arqueol. Etnol., Mendoza* **26**, 69 (1971).
5. Bahn, P.G. *Pyrenean Prehistory* (Aris & Phillips, Warminster, 1984).
6. Breuil, H. *Bull. Soc. préhist. fr.* **46**, 342 (1949).
7. Breuil, H. *Bull. Soc. Hist. nat. Toulouse* **90**, 245 (1955).
8. Cook, A.B. *L'Anthropologie* **14**, 655 (1903).
9. Lang, A. *Man* **4**, 37 (1904).
10. Piette, E. *L'Anthropologie* **7**, 385 (1896); **14**, 641 (1903).
11. Marshack, A. *Valcomonica Symp.*, 479 (1970).
12. Thévinin, A. *19th Congr. Préhist. fr.*, 341 (1972); *Gallia Préhist.*, 26 (1983).
13. Couraud, C. *La Recherche* **109**, 356 (1980).
14. Couraud, C. & Bahn, P.G. *Proc. prehist. Soc.* **48**, 45 (1982).

© 1984 Nature Publishing Group

Semen and AIDS

SIR — Since the demonstration that sperm can be immunosuppressive[1], and the postulation that exposure to semen by rectal intercourse could be a factor in the aetiology of acquired immune deficiency syndrome (AIDS)[1-3], some discussion and confusion have resulted concerning the exposure of men and women to semen[4,5].

Specifically, the issue deals with why homosexual men who are exposed to semen by anal intercourse would be immunosuppressed, and thereby more susceptible to infections, whereas women who are also exposed to semen during vaginal intercourse would not appear to be immune-compromised. This apparent contradiction can be accounted for if one compares the tissue structure of the vagina and the rectum[6]. The vagina (Fig.1) is lined by thick stratified squamous epithelium. This multilayered epithelial lining makes ulceration and penetration of semen into the underlying vascular lamina propria unlikely during normal sexual intercourse. In contrast, the rectum (Fig. 2) is lined by simple columnar epithelium. A single layer of cells (which can be easily ulcerated

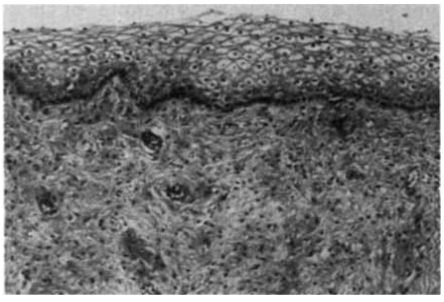


Fig. 1 Section of vaginal mucosa;

during anal intercourse) separates the semen from blood vessels and lymphatics of the rectal lamina propria. Thus, the rectum is more susceptible than the vagina to mucosal ulceration, and exposure of the vascular system and draining lymphatics to semen is more likely to occur.

Although some women engage in anal intercourse, it would be exceptional for women to practise anal intercourse as frequently as promiscuous homosexual men who get AIDS. Since a number of women have contracted AIDS[7], an important aetiological factor could be whether some of these patients have practised anal intercourse, particularly with bisexual men. In this context, it is also possible that the semen of some men who engage in homosexual practices may have greater immunosuppressive potential than semen of heterosexual men, possibly due to autoimmune-induced changes in cellular and/or soluble components of semen. If this is true, women who engage in anal intercourse with bisexual men may be more likely to be immunosuppressed than those who have anal sex with heterosexual men.

It should be noted in this context that a recent study has demonstrated that a

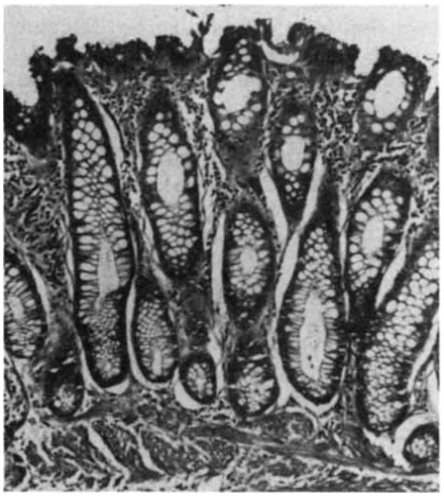


Fig. 2 Section of rectal mucosa.

woman who routinely engages in anal intercourse exhibited immune abnormalities[8]. Although the differences in the epithelium of the rectum and vagina can account for differences in susceptibility to semen-induced immunosuppression, the possibility is not excluded that females may be naturally more resistant to such suppression than males.

G. M. SHEARER
A. S. RABSON

Division of Cancer Biology and Diagnosis, National Cancer Institute, Bethesda, Maryland 20205, USA

1. Hurtenbach, U. & Shearer, G.M. *J. exp. Med.* **155**, 1719 (1982).
2. Shearer, G.M. & Hurtenbach, U. *Immun. Today* **3**, 153 (1983).
3. Sonnabend, J., Witkin, S.S. & Purtilo, D.T. *J. Am. med. Ass.* **249**, 2370 (1983).
4. De Tolla, U. *Nature* **304**, 678 (1983).
5. McLaren, A. *Nature* **304**, 678 (1983).
6. Bergman, R.A. & Afifi, A.K. *Atlas of Microscopic Anatomy* (Saunders, Philadelphia, 1974).
7. *MMWR CDC Update:* **32**, 465 (1983).
8. Mavligit, G.M. *et al. J. Am. med. Ass.* (in the press).

Evolutionary unease

SIR — Evolution biology has the curious distinction that although it has an extensive literature of cogent observation and relevant theory, and in this sense is among the best established branches of science, it periodically suffers attacks, as it recently has in the popular news media, from critics to whom the subject appears to have some emotional or supposedly religious significance. Such critics often address themselves to what they call "Darwinism" although internal evidence makes it doubtful whether they have in fact read Charles Darwin's *Origin of Species*, let alone the modern superstructure of genetics and molecular biology that has been built on the foundation he laid. Indeed some fundamentalists seem not to discuss the distinction between the "P" creation story beginning in Genesis 1.1, and the "J" account in Genesis 2.4.

More professional criticism has come from Sir Fred Hoyle and N.C. Wickramasinghe (*Evolution from Space*, Dent, London, 1981) who show that the probability of a viable organism arising by chance is hopelessly low if the known and powerful effect of multistage selection is omitted from the statistical calculation. This is of course entirely as expected, and no evidence against received evolutionary theory. Other critics have interpreted the rarity of intermediate types in the incomplete geological record without due regard to modern mathematical catastrophe theory.

While there is thus little to suggest that any major revision is needed to current theories, a degree of unease remains. This gives rise to various *ad hoc* suggestions, of which one of the latest is "molecular drive". Suggestions of this nature seem unnecessary if full account is taken of the cybernetics of evolution. Darwin's recognition that hereditable characteristics favouring survival will be more likely to survive is effectively tautological, and as such is unassailable. Since offspring both resemble and differ from their parents, it follows logically that evolution must proceed *at least* by natural selection, and modern genetic theory and molecular biology have elucidated in some detail how this happens. This does not exclude the possibility of additional mechanisms, but such should be invoked only if their necessity were demonstrated.

In assessing whether the power of natural selection is alone sufficient, it is necessary to take account of the cybernetics of multiple feedback loops. There will be selection not only of organisms by and for survival, but also selection of the way in which genetic material is held. Groups of organisms will be favoured which code their genetic information so that mutations are more likely to result in viable individuals, some of which may have enhanced prospects for survival. Put simply, there is selection for selectability, giving rise to a double feedback-loop cybernetic system with enhanced power to stabilize the path of evolution along favourable lines. This capability may be further increased by additional orders of selection for selectability, continuing by mathematic induction to any depth without limit.

The physical mechanisms and quantitative effectiveness of this higher-order selection remain to be elucidated, but in computer applications it is known that controlled redundancy, distributed commands and related techniques may be employed so to structure a program that local changes tend to produce a new program that runs rather than one that just crashes. There are suggestive resemblances between these techniques and what is known of the distributed, redundant, molecular nature of the structures now identified with functional genes.

PETER FELLGETT

Department of Cybernetics, The University of Reading, Reading RG6 ZAL, UK

© 1984 Nature Publishing Group

Phosphorus, phosphorites and skeletal evolution at the Precambrian–Cambrian boundary

Peter J. Cook & John H. Shergold

Division of Continental Geology, Bureau of Mineral Resources, PO Box 378, Canberra, ACT 2601, Australia

Clustering of sedimentary phosphate deposits around the Precambrian–Cambrian boundary is related to a global phosphogenic event. This event probably resulted from a complex interplay of factors and had a profound effect on the biota, with high concentrations of phosphorus in the photic zone producing an increase in the total shallow marine biomass, faunal diversification, the precipitation of skeletal structures and explosive evolution.

THE change from soft bodied to shelly faunas which took place at the Precambrian–Cambrian (Vendian–Tommotian) boundary is one of the most dramatic, important, and least understood events in the geological record. The reason for this event, often referred to as the Cambrian explosion or Cambrian radiation event[1], has been the subject of much speculation. Possible explanations for it include the attainment of a critical atmospheric oxygen level[2] or a decrease in the concentration of atmospheric carbon dioxide[3], the availability of new shallow-water ecological niches as a result of a worldwide marine transgression[4,5] and an enhanced availability of food (especially phytoplankton) related to a major event of oceanic upwelling[6,7]. These various suggestions are by no means mutually exclusive. Here we will attempt to relate these evolutionary changes to the distribution of phosphorites around the Precambrian–Cambrian boundary and to the global events which may have been responsible for that distribution.

Phosphorites

Phosphorites, phosphatic sedimentary rocks, composed of abundant collophane (usually cryptocrystalline carbonate-fluorapatite) form major phosphate deposits in many parts of the world and occur throughout the geological column, though especially within the Phanerozoic. In the 150 yr since their discovery, various mechanisms have been proposed to account for the formation of phosphorites[8–10]. The best known hypothesis is undoubtedly the so-called 'upwelling hypothesis' first proposed by Kazakov[11], who suggested that phosphorites formed as a result of the upwelling of deep cold low *p*H oceanic waters onto the shelf, and that this led to the essentially inorganic precipitation of phosphate (as apatite) from the water column. This deceptively simple theory appeared to explain many of the observed features of phosphorites such as those in the Phosphoria Formation[12]. However, in recent years it has become increasingly apparent that the original upwelling hypothesis does not in fact offer an adequate explanation. For example, many deposits can be quite clearly demonstrated to have formed (or accumulated) under very shallow marine conditions and there is abundant evidence of the importance of diagenetic phosphatization. Also, major geochemical problems are encountered in precipitating phosphate from seawater. As a consequence of these problems it has become recognized that phosphogenesis is a complex multistage process. For example, grainstone phosphorites—the most common type of deposit— probably form as a result of a series of biochemical and sedimentary processes involving: (1) influx of nutrient-rich waters into or adjacent to a very shallow marine region; (2) development of a prolific biota; (3) formation of anoxic organic-rich bottom sediments; (4) leaching of phosphorus from organic-rich sediments leading to some localized precipitation of phosphate or more commonly phosphatization of the associated sediments; (5) reworking of the sediments to produce concentrations of phosphate grains; (6) shoreward transport and entrapment of phosphate grains in shallow nearshore zones to form high grade phosphate deposits.

This genetic scheme represents a marked departure in detail from the original upwelling hypothesis of Kazakov but it should be noted that an important element may persist in the initial step to produce a high organic productivity system. Areas of abnormally high organic productivity can develop as a consequence of enhanced nutrient input from rivers or fumaroles for example, but there is no doubt that in the present day oceans there is an overwhelming preference for elevated phosphorus levels in the ocean to be associated with areas of upwelling. The palaeontology, geochemistry and palaeogeography of many ancient phosphorites, together with the formation of modern phosphorites in offshore Chile–Peru and Namibia, are consistent with a link between many (though not all) phosphorites and oceanic upwelling. It is reasonable to assume that this link is provided by the elevated phosphorus content of the waters and the consequent development of high organic productivity. However, contrary to earlier hypotheses directly linking upwelling to the formation of phosphorites, phosphate deposits may have accumulated far from the actual site of upwelling. Many of the major deposits accumulated within very shallow epicontinental seas with water depths of generally no more than tens of metres, sites where oceanic upwelling could not have occurred.

Detailed consideration of the palaeolatitudes of phosphorites led Cook and McElhinny[13] to develop models for phosphogenesis related to low latitude locations, rifting and particular phases of seafloor spreading. The latitudinal model resulted from the development of a narrow east–west oriented near-equatorial seaway (such as the Tethys, which resulted in the major Cretaceous–Eocene phosphate deposits of North Africa). The longitudinal model was the consequence of the formation of a broad north–south seaway (such as the Palaeo-Pacific, which resulted in for instance, the Permian Phosphoria deposits). These two basic models, together with a third involving the development of monsoonal upwelling, are used by Zeigler and co-workers[14,15] to test proposed Phanerozoic reconstructions and climates. Arthur and Jenkyns[16] have shown for the Mesozoic and Cenozoic that 'phosphorite giants' correlate in a general way with elevated sea level and climate, and in turn with plate tectonics and continental drift. Oceanic anoxic events (OAEs)[17] do not coincide with global phosphogenesis[16], though locally, phosphorites and organic-rich sediments may be associated.

© 1984Nature Publishing Group

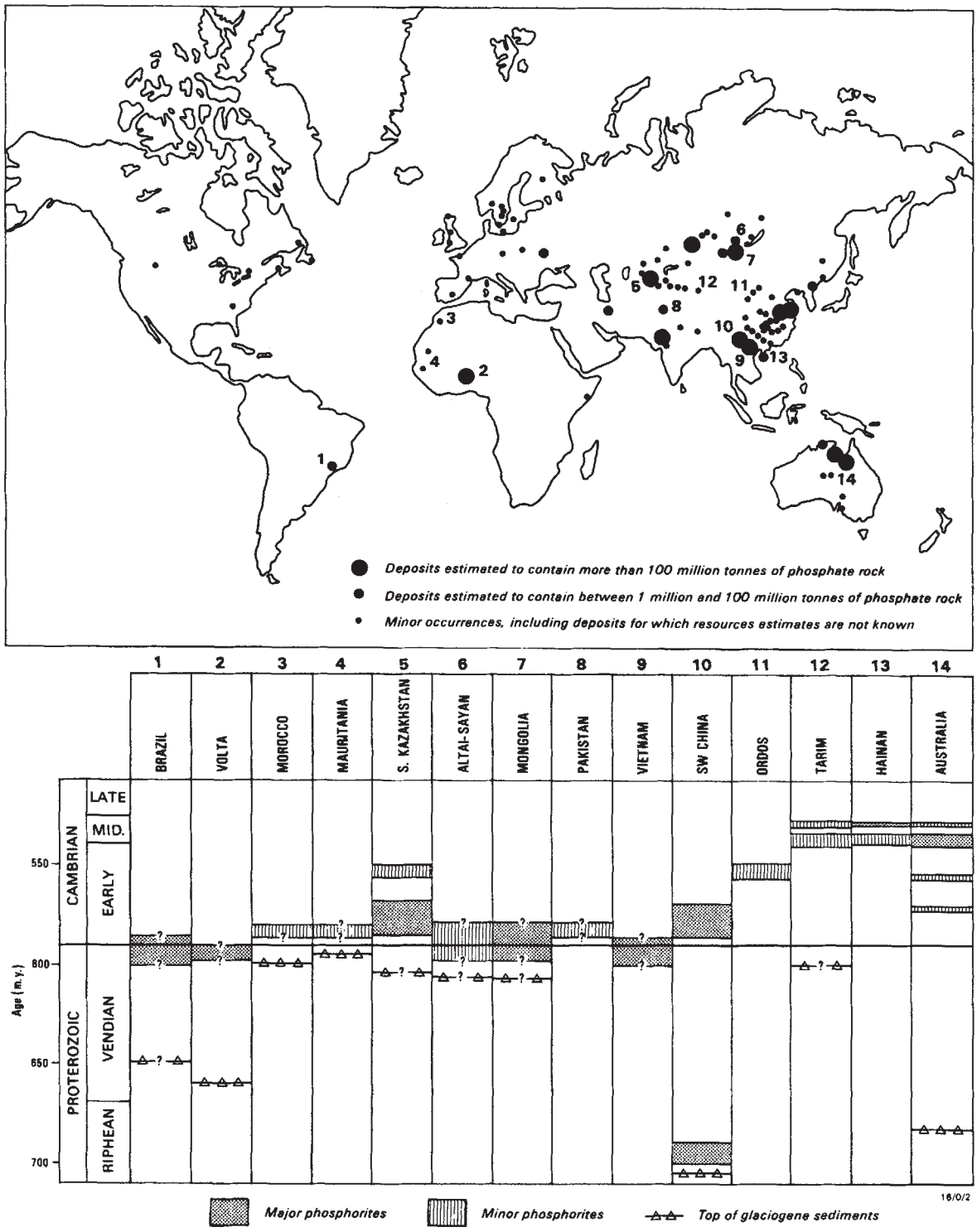


Fig. 1 Spatial and temporal distribution of Proterozoic and Cambrian phosphorites. The geochronological scale on the left is adapted directly from Harland *et al.*[50], as are the chronostratigraphic boundaries. The distribution and time range of the glacial and tillitic deposits indicated is based largely on various authors in Hambrey and Harland[31] with the exception of column 10 where Compston and Zhang[51] are followed. Authorities consulted for the stratigraphic positioning of the phosphorites are for columns 1–14: (1) Cathcart[30], Borrello[52]; (2) Trompette *et al.*[28], Affaton[27]; (3) Viland[53], Bertrand-Sarfati[54]; (4) Poulsen[55], Bassot[56], Sougy[57,58]; (5) Rozanov[59], Eganov[60], Missarzhevsky and Mambetov[61]; (6) Kazarinov and Krasilnikova[62]; (7) Bjamba[63], Ilyin[24], Ilyin and Bjamba[25]; (8) Bhatti *et al.*[26], Hasan *et al.*[64], Shah[65]; (9) Kalmykov *et al.*[21]; (10), (11), Lu Yanhao[22], Wang Yangeng[32], Luo Huilin *et al.*[66]; (12), (13), Lu Yanhao[22], Chang Wentang[67]; (14) Callen[68], Fleming[69], Knight[70].

© 1984Nature Publishing Group

There is clear evidence that certain periods of Earth history have been particularly favourable for the formation of phosphorites, notably Miocene–Pliocene, late Cretaceous–Eocene, Jurassic, Permian, middle Ordovician and late Proterozoic–Cambrian times. There are a number of older Proterozoic episodes, but they are less important and the deposits formed are much smaller. The late Proterozoic–Cambrian phosphogenic period appears to be the first major episode, with phosphorite deposits and occurrences in many parts of the world (Fig. 1), though particularly in the central and South-East Asian region and Australia[18–20]. Major deposits containing many billions of tonnes of phosphate rock occur in the Georgina Basin of northern Australia[18], the Lao Cai area of Vietnam[21], southern and central China[22], the Karatau area of Kazakhstan[23], and the Khubsugul Basin of Mongolia[24,25]. Lesser known deposits also occur in the Hazara region of north-west Pakistan[26] and Soltanai, Iran. In Africa, there are at least two late Proterozoic–Cambrian phosphogenic regions, one extending through neighbouring parts of Ghana, Nigeria, Niger, Upper Volta, Dahomey and Mauritania and the second, lesser known, in Somalia[27–29]. In South America, deposits also occur in the Bambui Group of Brazil[30]. This interval of time is also notable not only for major deposits but also for abundant phosphorite occurrences and phosphatic sediments. These are particularly widespread in the Cambrian sedimentary sequences of the Baltic and Avalon Platforms of Europe and maritime eastern North America, the Altai–Sayan folded region of southern USSR, the Siberian Platform, and in southern and central Australia[19]. To date, Antarctica is the only continent in which there are no known latest Proterozoic or Cambrian phosphatic sediments. In all, more than 100 important phosphorite deposits and occurrences have been documented and phosphate resources of this age are probably second only to those in the late Cretaceous and Eocene in terms of future economic importance. This great economic significance is all the more remarkable when it is remembered that probably little of the original quantity of upper Proterozoic and Palaeozoic sediments has been preserved (J. Veizer, personal communication suggests approximately 30% remains).

While the age of many late Proterozoic–Cambrian deposits and occurrences cannot be given with great precision, if all available biochronostratigraphic data are assembled for major deposits (Fig. 1), it is apparent that the great majority are of early Cambrian age. Even major deposits, such as those of the Khubsugul Basin of Mongolia, the Guiyang deposits of China or some of the deposits of West Africa, which are generally regarded as late Proterozoic (because of the absence of macrofossils), cannot be accorded a Precambrian age with any certainty. All these inferred Precambrian deposits are located above latest Precambrian tillites (which suggests an age of less than about 650 Myr)[31] and below sediments containing the earliest Cambrian shelly fossils. Consequently, some of these late Proterozoic deposits could in fact be located at, or even immediately above, the Precambrian–Cambrian boundary. This suggests a single widespread very late Proterozoic to early Cambrian phosphogenic episode commencing no earlier than the Riphean–Vendian boundary and probably much closer to the base of the Cambrian, reaching a peak during the early Cambrian (between about 590–550 Myr), then rapidly decreasing during the earliest part of the middle Cambrian before finally terminating during the late middle Cambrian at about 540 Myr. Alternatively, two distinct phosphogenic episodes may have occurred—one in the very late Proterozoic and the second, and most important, in the early Cambrian. For example, the major deposits of China occur in the late Proterozoic Doushantuo Formation of Guizhou Province and the early Cambrian Meishucun Formation of Yunnan Province, suggesting a hiatus. But in fact the widespread dolomitic Denying Formation, which occurs between these two phosphatic intervals, whilst containing no major phosphorite deposits, is found to be phosphatic, so that phosphogenesis may have been continuous from the very latest Proterozoic into the early Cambrian[22,32]. Therefore the first option of a single phosphogenic episode spanning the Precambrian–Cambrian boundary is favoured because of the lack of any real evidence of a phosphogenic hiatus at this time, though there may be smaller sub-episodes within this major one.

Although it is possible to demonstrate that Cambrian phosphate deposits do in general have low palaeolatitudes[13], present palaeomagnetic data for latest Proterozoic and Cambrian time are inadequate for anything more than the most generalized global reconstructions. Nevertheless, present evidence suggests that all the major deposits formed in shallow epicontinental lobes off a major seaway at low palaeolatitudes. Additionally, all the major deposits occur in a strikingly similar sedimentary sequence. This is not to suggest that all the deposits are synchronous, but rather that the same basic sedimentological and geochemical processes have been operative for all the deposits. As mentioned earlier, many of the processes leading to the formation of phosphate deposits are local in nature, for example the right configuration of coastline to produce a phosphate 'trap'. Nevertheless, the occurrence of a large number of deposits all of about the same age, in many parts of the world (Fig. 1), suggests a basic underlying cause. The fortuitous location of many of the shelf seas at low latitude locations could be one such cause, sea-level changes a second, whilst changes in the composition of the world ocean could be a third. Although 'phosphate giants' may not correlate with OAEs[16] in fact, during the Mesozoic and Cenozoic, the only time for which there are adequate data, the major phosphogenic events in the Miocene–Pliocene, late Cretaceous–Eocene and middle–late Jurassic are each preceded by OAEs in the middle Miocene, the early–middle Cretaceous and the early–middle Jurassic respectively. The present data on OAEs for the late Proterozoic and early Palaeozoic are inadequate to allow us confidently to make the same correlations.

Geochemical evidence

There are however some lines of evidence, notably stable isotopic data, which suggest that the same relationship (of an OAE preceding phosphogenesis) may have applied. $\delta^{34}S$ values for late Proterozoic and early Palaeozoic evaporites indicate major changes in the isotopic signature at this time[33,34]. The so-called Yudomski event which occurs just below the Precambrian–Cambrian boundary[33] is particularly marked though the timing of it is still somewhat imprecise. This event involves a change in the sulphur isotope ratio of evaporitic sulphate from about +16‰ to +31‰, a $\delta^{34}S$ change of 15‰. Although there are other possible scenarios, the one favoured here is similar to that proposed by Holser[34], with this major isotopic excursion resulting from 'catastrophic' ocean water mixing, except that there the mixing is of deep and shallow oceanic waters rather than of oceanic water and water from evaporitic basins.

Few sulphur isotope values are available for lattice sulphate in the apatite of phosphorites, but those available (Fig. 2) appear to be consistent with the $\delta^{34}S$ values for sulphate evaporites[35]. These preliminary data suggest that a major increase in the $\delta^{34}S$ of phosphorite sulphate seems to occur around the Precambrian–Cambrian boundary (Fig. 2) though the data are admittedly sparse. The shift in the values of the sulphur isotopic masses between the deep and shallow marine reservoirs is also reflected in changes in $\delta^{13}C$ (ref. 36). A depletion in $\delta^{13}C$ of 2‰ occurs around the Precambrian–Cambrain boundary (Fig. 2). The data are insufficient to state confidently that this change occurs right at the boundary. Nevertheless, the correlation which exists between $\delta^{13}C$ (carbonates) and $\delta^{34}S$ (sulphate) (ref. 36) is consistent with our thesis that phosphogenesis around the Precambrian–Cambrian boundary was preceded by a widespread relatively anoxic phase and that this is reflected in both the sulphur and carbon isotopic values of the sediments.

The sequence of events prior to the Yudomski event, may have been that the rate of mixing of the deeper parts of the world ocean was abnormally low. With stagnation in places, anaerobic sulphate-reducing bacteria generated H_2S and pyrite,

© 1984Nature Publishing Group

Table 1 Some reservoirs and fluxes for phosphorus in the present-day ocean

R_b	(P reservoir marine biomass)	$=4.6\times10^{11}$ mol P
R_s	(P reservoir shallow ocean)	$=7.4\times10^{13}$ mol P
R_d	(P reservoir deep ocean)	$=2.4\times10^{15}$ mol P
F_p	(P flux shallow to deep ocean)	$=1.35\times10^{12}$ mol yr^{-1} P (particulate)
F_r	(P flux river to shallow ocean)	$=3\times10^{10}$ mol yr^{-1} P

Values from Sandstrom[40].

and the residual SO_4 became $\delta^{34}S$ enriched. The deep ocean water was characterized by isotopically 'heavy' sulphate and high phosphorus. The shallow ocean, which is that part of the water column generally 'sampled' in the geological record, would have been characterized by depletion of phosphorus and by a lighter $\delta^{34}S$ signature. This low-circulation (polytaxic) period[17] was followed by the Yudomski event, when there was much more vigorous overturn of the water column (an oligotaxic period[17]) and a consequent spread of high phosphorus–'heavy' sulphur water into the shallow marine zone. The main region of anoxia was now in the shallow water areas, especially where upwelling impinged on the continental shelf. However, anoxic, high phosphorus conditions also spread far across the shelf and into epicontinental seas in places, as indicated by the widespread occurrence of black organic shales (and the 'stone coals' of China) at about this time. The occurrence of black shales and phosphorites is in fact an extremely common association throughout the late Proterozoic and Cambrian.

Such compositional changes in the world oceans have been previously proposed[37–39], though what triggered the initial overturn of the stagnant ocean and the consequent spread of high phosphorus isotopically 'heavy' sulphate-containing waters into the photic zone is not clear. Possible causes include the formation of dense cold surface waters by the latest Precambrian glaciations, or dense saline surface waters by a period of prolonged surface evaporation. More probably, it was related to plate movement, and rifting which developed seaways favourable for oceanic upwelling, or the movement of coasts to preferred low latitude locations at about the time of the Precambrian–Cambrian transition. Widespread volcanism is consistent with this being a time when plate movement was initiated.

The available geochemical data are inadequate for modelling the present day global phosphorus cycle with any precision, let alone the phosphorus cycle for the Precambrian or Cambrian. If, however, some present day values are taken[40] (Table 1) then it is possible to show the possible effects, albeit as 'caricatures' of OAEs. Take for example the extreme situation where all phosphorus exchange via oceanic overturn ceased between the deep and shallow ocean. The shallow ocean reservoir would be depleted of almost all phosphorus within time

$$t=\frac{R_b}{F_p}=55\text{ yr}$$

This would in turn lead to an increase of 3% in the phosphate of the deep ocean reservoir. If it is assumed that the fluvial flux (F_r) was entirely used by the primary producers (and $F_r=F_p$) then the time taken to double the phosphorus content of the deep ocean would be

$$t=\frac{R_d}{F_r}=8\times10^4\text{ yr}$$

If all of the excess phosphorus (2.4×10^{15} mols P) was then circulated back in solution to the shallow ocean in such a way that it was not all used up by the biota (again unlikely in reality), then the total quantity of phosphorus in the shallow ocean would increase by $R_d/R_b=522$ times! The point of these extreme scenarios, and the admittedly outrageous geochemical budgets that they generate, is not to imply that the biomass did indeed increase by 522 times, but rather to make the point that a decrease in deep-shallow exchange even for a period of 10^4–10^5 yr could have a profound effect on phosphorus distribution, and that the deep ocean could become a major phosphorus ink. However, if circulation was resumed, then this could have an even more profound effect on the oxygen and CO_2 levels in the atmosphere if there was an explosion in the total biomass of photosynthesizers[41,42].

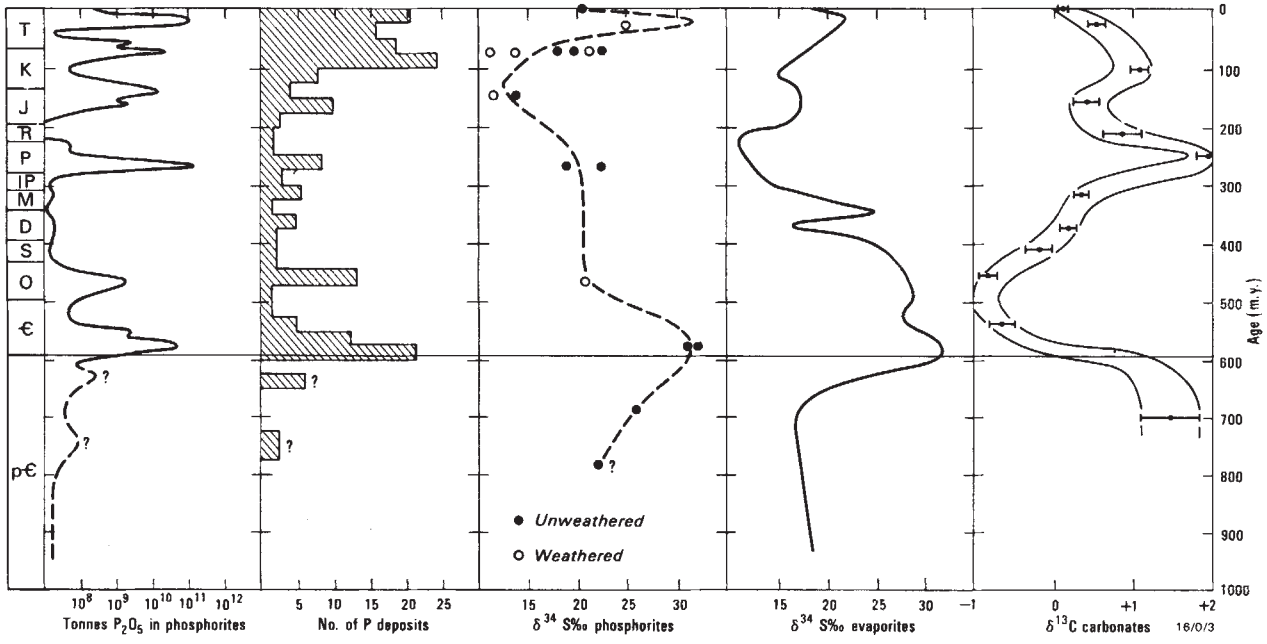


Fig. 2 Abundance of phosphorite deposits during the late Precambrian and Phanerozoic (modified after Cook and McElhinny[13]) and sulphur and carbon isotope curves for evaporite sulphates and carbonates respectively (after Claypool *et al.*[71] and Veizer *et al.*[36]). $\delta^{34}S$ values for lattice sulphate in phosphorites is after Blishovsky *et al.*[35].

© 1984Nature Publishing Group

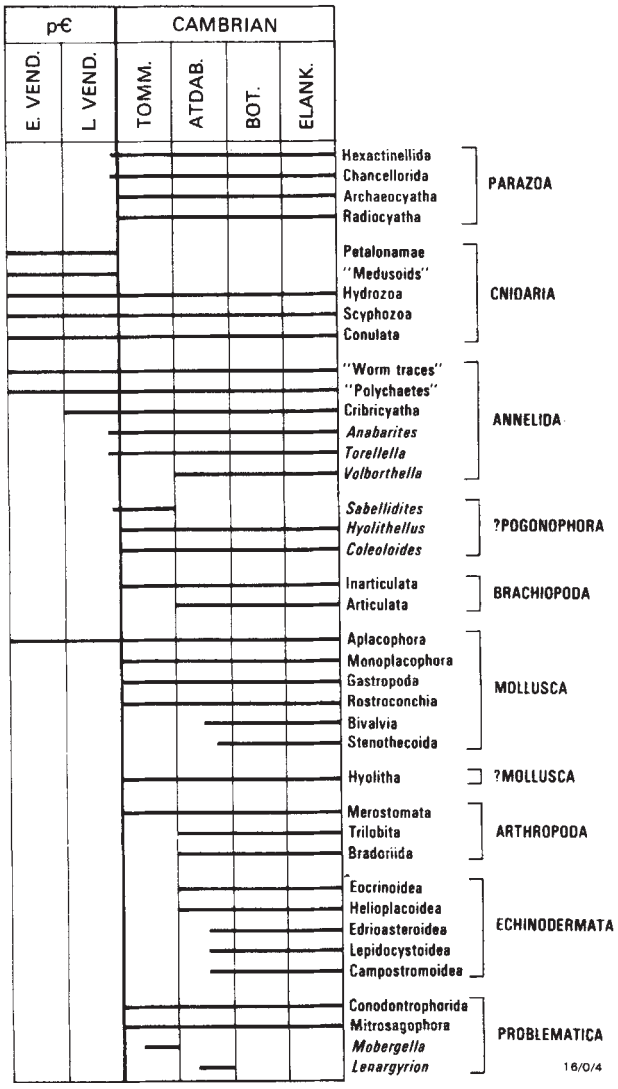


Fig. 3 Stratigraphic distribution of late Precambrian and Cambrian macrofossils (after Stanley [72], Runnegar and Pojeta[73], and Brasier[1]).

Therefore, changes in the rate of oceanic overturn around the Precambrian–Cambrian boundary could potentially have exerted a direct and dramatic effect on the biota at that time. In many parts of the present-day world ocean, plant and animal growth is phosphorus-limited[10]. The phosphorus distribution is primarily controlled by upwelling of deep phosphorus-rich oceanic waters and the movement of attendant oceanic currents. Upwelling and the associated distribution of phosphorus is of fundamental importance to the distribution of the present-day marine biota. It would be unreasonable to assume that distribution and availability of phosphorus was any less important in the past. This in turn suggests that synchroneity of a major phosphogenic event, a major evolutionary event, and the Precambrian–Cambrian boundary are likely to be the consequence of something more profound than chance?

The fossil record

Turning to the fossil record for the late Proterozoic and Cambrian (Fig. 3) we see that there are several evolutionary milestones[43–45]. These include the appearence of (1) soft-bodied metazoans (the Ediacara fauna) at about 670 Myr; (2) metazoan burrowers in the late Proterozoic; (3) mineralized shelly faunas close to the Precambrian–Cambrian boundary at about 595 Myr; (4) the first organic predators; and (5) a wider range of ecological strategies sought and used during the initial Cambrian.

Within the limits of available geochronology the first appearance of the metazoans corresponds reasonably well with the commencement of late Proterozoic phosphogenesis. Even more marked, is the coincidence of the main phase of phosphogenesis in the early Cambrian with the appearance of the first biomineralized skeletons. The first mineral skeletons to appear were composed almost solely of Ca-bearing compounds particularly carbonates and phosphates. Many hypotheses, summarized by Brasier[5], have been advanced to explain why biomineralization developed at this particular time. These include, adaptation for burrowing, refinement of Ca-ion regulation, and the development of the atmospheric O_2 (ref. 46) and its relationship to organic respiratory evolution.

The hypothesis followed here is that a major increase in the phosphorus content of the marine photic zone occurred at about the Precambrian–Cambrian boundary as a consequence of enhanced oceanic overturn, following a period of decreased overturn. This event was reflected in the widespread occurrence of phosphorites, and also provided the driving force for many of the evolutionary changes at the Precambrian–Cambrian boundary. Such an event would have resulted in a major increase in the shallow marine biomass, particularly the phytoplankton. This in turn could have produced an increase in the oxygen content of the atmosphere at this time, an event previously postulated by several authors. However, it is important to note that here it is proposed that the driving mechanism put forward is an increase in the nutrient levels of the marine photic zone, which resulted in an enhanced biomass and an attendant increase in free oxygen, as a result of an increased level of photosynthesis, rather than oxygen necessarily being the dominant control on evolution. It is germane to note that for much of the total biomass, oxygen is relatively unimportant or even toxic, whereas there is no form of life that can exist without phosphorus, a feature which perhaps has not received the attention to which it is entitled.

Abundance of nutrients is consistent with the great variety of body plans observed in the Cambrian fossil record. Brasier[1] points out that "this variety in the Cambrian may have exceeded that at any other time in Earth history, but many phyla and classes were geologically short-lived, excluded perhaps by the progressive race for efficiency". Thus, a situation is likely to have developed of rapid utilization of available environmental niches, and competition for them, in turn leading to a rapid increase in skeletal development. This permitted reorganization of visceral organs, nervous and vascular systems to cope with changing habitats, generating an evolutionary advantage. With progressive depletion of nutrient supply during the early Cambrian, as progressively more phosphorus was lost in the sedimentary 'sink', a complexity of trophic levels developed, together with an intensification of organic skeletal development. An exoskeleton could well have been acquired as a defensive mechanism during periods of high trophic competition, and/or may have given the owner an advantage in feeding and reproducing, or merely allowing progeny to survive longer. Alternatively it could have been a chance to excrete excess quantities of an unwanted element or compound.

A relative abundance of phosphorus available during the latest Precambrian and initial Cambrian is also reflected in the relatively high proportion of organisms with calcium phosphate rather than calcium carbonate skeletal structures. As pointed out by Lowenstam and Weiner[47] "Just prior to the Cambrian period, all the known biogenic minerals were calcium-minerals and two thirds were composed of calcium phosphate". By late in the early Cambrian, and throughout the remainder of the Phanerozoic, calcium carbonate and chitin became the main skeletal components. Primary phosphatic elements are clearly more abundant in the initial Cambrian Tommotian Stage than in succeeding stages. They are also more abundant in the vicinity of phosphate deposits, suggesting that the availability of phos-

© 1984Nature Publishing Group

phorus does indeed facilitate the development of calcium phosphate skeletal structures. Phosphorus also has an important secondary effect, namely as an inhibitor of calcification[48,49]. Although the precise mechanism is uncertain, it appears that inhibition results from surface 'poisoning' of calcium carbonate crystals by complex phosphate ions, especially orthophosphates[48]. Thus the development of calcareous skeletal structures is likely to have been inhibited until a significant decrease in the phosphorus concentration of the water column had taken place, presumably by the end of early Cambrian time. Calcareous hard tissues then became overwhelmingly dominant, a situation which persisted throughout the remainder of Phanerozoic time.

Conclusions

Many of the well-documented features at the Precambrian–Cambrian boundary can be related to a major phosphogenic event at about that time. A demonstrable sustained period of phosphogenesis commenced in the terminal late Proterozoic, reaching a maximum in the initial early Cambrian. This event was probably related to a period of enhanced oceanic overturn following a prolonged period of anoxia during which high-phosphorus waters were formed in the deep ocean. Whilst the reasons for this change are unclear, the final glaciation of the late Proterozoic may have been a possible trigger mechanism. Alternatively, and more probably, the change was induced by a phase of continental drift which provided an ocean–continent configuration (possibly a narrow latitudinal near-equatorial ocean) that produced an increased rate of oceanic overturn. An important component may also have been major sea-level rises and the formation of extensive shallow epicontinental seas, which were available for colonization and also provided 'traps' for phosphate grains. These changes are reflected in the isotopic signatures of the Yudomski Event. This, and the associated phosphogenic event was in turn followed by the so-called Cambrian radiation event which is recorded in the fossil record by the initiation of biomineralized skeletal tissues in various organisms all at about the same time. A significant proportion of these early skeletons was built of calcium phosphate, which may be interpreted as an opportunistic or alternatively a crisis reaction to the high levels of phosphorus in the water column of the photic zone. There may have been a massive increase in the biomass of the photic zone, particularly of the phytoplankton. It is possible that this increase in the biomass was sufficient to increase the free oxygen level significantly through enhanced photosynthesis at around the Precambrian–Cambrian boundary. As phosphorus was withdrawn progressively from the shallow seas into the phosphorus sink represented by phosphorites, the phosphorus content of the photic zone gradually decreased to the point where phosphorus was no longer so readily available for biomineralization. Additionally, it no longer constituted an inhibitor of calcite precipitation and as a consequence, calcitic and aragonitic skeletons then increased in abundance, to attain the preeminence which persisted throughout the remainder of the Phanerozoic.

This article is a contribution to the UNESCO-IUGS International Geological Correlation Program Project 156-Phosphorites. We thank Terry Donnelly, Mark Sandstrom, Dick Sheldon, Peter Southgate, Phil Trudinger, Jan Veizer and Sen Shu Sun for their comments on the manuscript. The article is published with the permission of the Director, Bureau of Mineral Resources.

Received 11 October; accepted 18 November 1983.

1. Brasier, M. D. *Systematics Ass. Spec. Vol.* **12,** 103–159 (Academic, London, 1979).
2. Cloud, P. E. *Science* **160,** 729–736 (1968).
3. Walker, J. C. G. *Evolution of the Atmosphere* (Macmillan, New York, 1977).
4. Matthews, S. C. & Cowie, J. W. *J. geol. Soc. Lond.* **136,** 133–135 (1979).
5. Brasier, M. D. *Precamb. Res.* **17,** 105–123 (1982).
6. Yanshin, A. L. in *Geology of Phosphorite Deposits and Problems of Phosphorite Genesis,* 13–14 (USSR Academy of Sciences, Novosibirsk, 1982).
7. Rhodes, F. H. T., & Bloxam, T. W. in *Proc. North Am. Paleont. Conv. Pt K,* 1485–1513 (Allen, Lawrence, Kansas, 1968).
8. Sheldon, R. P. *U.S. geol. Surv. Prof. Pap.* **501**C, 106–113 (1964).
9. Baturin, G. N. *Nature phys. Sci.* **232,** 61–62 (1971).
10. Baturin, G. N. *Phosphorites on the Sea Floor, Origin, Composition and Distribution* (Elsevier, Amsterdam, 1982).
11. Kazakov, A. V. *Trans. Sci. Inst. Fertil. Insecto-fungicides* **142,** 95–113 (1937).
12. McKelvey, V. E., Swanson, R. W. & Sheldon, R. P. *Int. geol. Congr. 19th Sess. Algiers* **11,** 45–65 (1953).
13. Cook, P. J., & McElhinny, M. W. *Econ. Geol.* **74,** 315–330 (1979).
14. Ziegler, A. M., Parrish, J. & Humphreville, R. in *Proterozoic-Cambrian Phosphorites* (eds Cook, P. J. & Shergold, J. H.) 21 (Australian National University Press, Canberra, 1979).
15. Ziegler, A. M., Scotese, C. R., MacKerrow, W. S., Johnson, M. E. & Bambach, R. K. *A. Rev. Earth planet. Sci.* 473–502 (1979).
16. Arthur, M. A. & Jenkyns, H. C. *Oceanol. Acta,* **45,** 83–96 (1981).
17. Fischer, A. G. & Arthur, M. A. *SEPM Spec. Publ.* **25,** 19–50 (1977).
18. *Proterozoic-Cambrian Phosphorites* (eds Cook, P. J. & Shergold, J. H.) 13 (Australian National University Press, Canberra, 1979).
19. Notholt, A. J. G. in *Proterozoic-Cambrian Phosphorites* (eds Cook, P. J. & Shergold, J. H.) 71–86 (Australian National University Press, Canberra, 1979).
20. Bushinski, G. E. *Old Phosphorites of Asia and their Genesis* (Israel Program for Scientific Translation, Jerusalem, 1969).
21. Kalmykov, A. F., Sokolov, A. S. & Tushina, A. M. *Trans-State Sci. Res. Inst., Min. & Chem.* **7,** 139–191 (1962).
22. Lu Yenhao *Cambrian Mineral Deposits in China and the Bio-Environmental Control Theory* (Geological Press, Peking, 1979).
23. Eganov, E. A. & Sovietov, Yu. I. *Karatau—a Model for Phosphorite Deposition* (Nauka, Novosibirsk, 1979).
24. Ilyin, A. V. *Joint Soviet-Mongolian Scientific-Research Geological Expedition Trans.* Vol. 6 (Nauka, Moscow, 1973).
25. Ilyin, A. V. & Bjamba, Z. *Guidebook for Excursion to Phosphorites of the Khubsugul Basin, Mongolian People's Republic, 17–28 June, 1980* (Soviet Committee for IGCP, Geological Institute, Academy of Sciences, Moscow, 1980).
26. Bhatti, N. A., Hussain, M. T. & Hussain, M. *Geol. Surv. Pakistan Inf. Release* **51,** (1972).
27. Affaton, P. *Trav. Lab. Sci. Terre St. Jerome, Marseille* **10,** 1–201 (1975).
28. Trompette, R., Affaton, P., Joulia, F. & Marchand, J. *Econ. Geol.* **75,** 62–70 (1980).
29. Ilyin, A. V. *Int. Geol. Correlation Program, Project 156 (Phosphorites), Newsl.* **2,** 8–9 (1978).
30. Cathcart, J. B. *Trans. Soc. Amer. Inst. Min. Engr. AIME* **264,** 1703–1708 (1979).
31. *Earth's Pre-Pleistocene Glacial Record* (eds Hambrey, M. J. & Harland, W. B.) (Cambridge University Press, 1981).
32. Wang Yangeng *The Phosphate-bearing Strata of the Late Precambrian to Cambrian in Upper Yangtze Region, China* (Geological Bureau of Guizhou, Huishui, China, 1982).
33. Holser, W. T., & Kaplan, I. R. *Chem. Geol.* **1,** 160–172 (1965).
34. Holser, W. T. *Nature* **267,** 403–408 (1977).
35. Blishovsky, V. Z., Grinenko, V. A., Migdisov, A. A., Savina, L. I. & Vernadsky, V. I. *Geochem. Int.* **14,** 148–155 (1973).
36. Veizer, J., Holser, W. T., & Wilgus, C. K. *Geochim. cosmochim. Acta* **44,** 579–587 (1980).
37. Holland, H. D. *Geochim. cosmochim. Acta* **36,** 637–657 (1972).
38. Degens, E. T. & Stoffers, P. *Nature* **263,** 22–27 (1976).
39. Berry, W. B. N., & Wilde, P. *Am. J. Sci.* **278,** 257–275 (1978).
40. Sandstrom, M. W. in *The Cycling of Carbon, Nitrogen, Sulfur and Phosphorus in Terrestrial and Aquatic Ecosystems* (eds Galbally, I. W. & Freney, J.) 133–142 (Australian Academy of Sciences, Canberra, 1982).
41. Cloud, P. E. *Am. J. Sci.* **272,** 537–548 (1972).
42. Cloud, P. E. *Paleobiology* **2,** 351–387 (1976).
43. Lowenstam, H. A. *Precamb. Res.* **11,** 89–91 (1980).
44. Lowenstam, H. A. & Margulis, L. *Biosystems* **12,** 27–41 (1980).
45. Cloud, P. E. & Glaessner, M. F. *Science* **217**(4562), 783–792 (1982).
46. Runnegar, B. N. *J. geol. Soc. Aust.* **29,** 395–411 (1982).
47. Lowenstam, H. A., & Weiner, S. in *Biomineralisation and Biological Metal Accumulation* (eds Westbrock, P. & de Jong, E. W.), 191–203 (Reidel, Dordrecht, 1983).
48. Simkiss, K. *Biol. Rev.* **39,** 487–505 (1964).
49. Black, M. *Palaeontology* **11,** 793–813 (1961).
50. Harland, W. B. *et al. A Geologic Time Scale* (Cambridge University Press, 1982).
51. Compston, W. & Zhang Zi-Chau in *Abstr. 6th Aust. Geol. Convn. Canberra 1983* Vol. 9, 235 (1983).
52. Borello, A. V. in *Lower Palaeozoic Rocks of the World* Vol. 1 (ed. Holland, C. H.) 385–438 (London, 1971).
53. Viland, J. C. *Notes Mém. Serv. Mines Carte géol. Maroc.* 13–22 (1977).
54. Bertrand-Sarfati, J. *Newsl. Stratigraphy* **10,** 20–26 (1981).
55. Poulsen, C. *Mat. Fys. Meddr. K. Danske, Videnskab. Selsk* **32**(7), 3–12 (1960).
56. Bassot, J. P. *Mém. BRGM* **40,** 1–322 (1966).
57. Sougy, J. *Am. Fac. Sci. Univ. Dakar* **15** (12), 1–695 (1964).
58. Sougy, J., *Notes Mém. Serv. geol. Maroc* **236,** 95–103 (1972).
59. Rozanov, A. Yu. in *Boundaries of Geological Systems* (ed. Peive, A. V.) 31–53 (Nauka Publishing House, Moscow, 1976).
60. Eganov, E. A. in *Proterozoic-Cambrian Phosphorites* (eds Cook, P. J. & Shergold, J. H.) 22–25 (Australian National University Press, Canberra, 1979).
61. Missarzhevsky, V. V. & Mambetov, A. M. *USSR Acad. Sci. Trans. geol. Inst.* **326,** 1–90 (1982).
62. Phosphorites and apatites of Siberia (eds Kazarinov, V. P. & Krasilnikova, N. A.) 1–232 (*USSR Acad. Sci. Siberian Division, Inst. Geol. Geophys, Novosibirsk* 1980).
63. Bjamba, Z. *Kwart. geol.* **17**(4), 853–877 (1973).
64. Hasan, M. T., Ghaznavi, M. I. & Hayat, A. S. *Geol. Surv. Pakist. Inf. Release* **74,** (1974).
65. Shah, S. M. I. (ed.) *Mem. geol. Surv. Pakist.* **12,** (1977).
66. Luo Huilin, Jiang Zhiwen, Zu Zhongjiu, Song Xuiliang & Zue Ziaofeng *Acta geol. sinica* **2,** 95–111 (1980).
67. Chang, W. T. *J. geol. Soc. Aust.* **27,** 137–150 (1980).
68. Callen, R. A. *Min. Resour. Rev. S. Aust.* **130,** 9–34, 80–94, 110–120 (1969).
69. Fleming, P. J. G. *Publ. geol. Surv. Qld* **3641,** (1977).
70. Knight, C. L. (ed.) *Economic Geology of Australia and Papua New Guinea,* 4 (Monogr. 4, Australasian Inst. Miner. Metall., Parkville, Victoria 1976).
71. Claypool, G. E., Holser, W. T., Kaplan, I. R., Sakai, H. & Zak, I. *Chem. Geol.* **28,** 199–259 (1980).
72. Stanley, S. M. *Am. J. Sci.* **276,** 56–76 (1975).
73. Runnegar, B. N. & Pojeta, J. *Science* **186,** 311–317.

© 1984Nature Publishing Group

Identifier sequences are transcribed specifically in brain

J. Gregor Sutcliffe, Robert J. Milner, Joel M. Gottesfeld & Richard A. Lerner

Research Institute of Scripps Clinic, 10666 North Torrey Pines Road, La Jolla, California 92037, USA

'Identifier' or ID sequences are present in 62% of the RNA polymerase II and III transcripts made in vitro *from brain nuclei but in fewer than 4% of the transcripts made from the nuclei of other tissues. An homologous 160-nucleotide cytoplasmic poly(A)$^+$ RNA species, BC1, and a smaller species, BC2, are located* in vivo *exclusively in neural tissues. Cloned ID sequences are polymerase III templates* in vitro. *Our data suggest a model in which brain-specific polymerase III transcription of ID sequences located in introns of brain genes activates those genes in a primary manner for polymerase II transcription.*

WE recently described a novel repeated rat nucleotide sequence which hybridized to an apparently brain-specific, small RNA[1]. In characterizing the messenger RNA of rat brain by studying cloned cDNA copies of cytoplasmic poly(A)$^+$ RNA, we discovered that a few (3%) of our cDNA clones containing inserts of 500–1,250-nucleotide pairs hybridized to an abundant 160-nucleotide RNA molecule which was present in brain, but not liver or kidney. Each of these clones contained a common 82-nucleotide sequence followed by an oligo(dA) tract (see Fig. 3*a*). The 82-nucleotide sequence was also found in an intron of the rat growth hormone gene[2], which is expressed in certain brain cells[3], as well as in the pituitary. We have termed the 82-nucleotide sequence an 'identifier' or ID sequence and reasoned that the sequence might be present specifically in brain genes, possibly in introns or other noncoding regions, to implement their tissue-specific expression, acting at either the level of transcription or mRNA maturation. The brain-specific 160-nucleotide RNA molecule is, then, either a by-product of the splicing of brain-specific mRNA or a brain-specific transcript *per se*.

This reasoning leads to several predictions. Two predictions are that genes which are uniquely expressed in brain will contain ID sequences and that there will be enough copies of ID sequences in the rat genome to account for a substantial portion of brain-specific gene expression. These predictions are borne out as described by Milner *et al.*,[4] who show that there are about 10^5 ID sequences in the rat genome, many of which are located in the introns of brain-specific genes. Another prediction is that brain primary transcripts, when viewed as a population, should be rich in ID sequences. Whether ID sequences are also highly represented in the transcripts of other tissues depends on whether ID sequences have any function or on the level at which ID sequences function (transcription or mRNA processing).

The present experiments represent our studies on the localization and transcription of RNA molecules containing ID sequences. We find that the 160-nucleotide species, which we call BC1 to discriminate it from ID, is localized exclusively in the cytoplasm of brain, pituitary and peripheral nervous cells, and that brain but not liver or kidney nuclei act as templates for both RNA polymerase II (Pol II) and RNA polymerase III (Pol III) for the transcription of RNA molecules containing ID sequences. These data suggest that brain-specific Pol III transcription of ID sequences might somehow regulate transcription by Pol II to allow the expression of neural-specific genes.

Specificity of BC1

We isolated total RNA from several rat tissues (adrenal, spleen, testes, heart, lung, muscle, intestine, thymus, pituitary and solar nerve plexus), fractionated it by gel electrophoresis and Northern blotted[5] to nitrocellulose. The blots were hybridized with nick-translated[6] ID-containing clones p1B224 or p2A120[1], thereby extending our original study which examined the RNA from only brain, liver and kidney. The ID probes hybridize to a 160-nucleotide target found only in brain, pituitary and solar plexus (Fig. 1). Our pituitary preparation contained both the posterior lobe, which is of neural origin, and the anterior lobe, which is an endocrine tissue. The hybridization to the RNA from each tissue by p1B15 (a cDNA clone of a ubiquitous rat mRNA) acts as a positive control to show that the various ID-negative lanes do not result from degraded RNA. Of particular note is that adrenal, which is partially derived from neural crest, and the rat adrenal phaeochromocytoma tumour PC12 were both negative for the 160-nucleotide RNA species in two separate experiments. Some of the samples used in the experiment of Fig. 1 were total cell RNA extracts: the small RNA is therefore absent from all compartments of the non-hybridizing tissues.

BC1 RNA

The ID-containing cDNA clones we have examined probably correspond to mRNA precursors contaminating our cytoplasmic RNA preparations because these clones contain both exon and intron sequences[4]. Therefore, the possibility existed that the small RNA was predominantly nuclear. We prepared nuclear and cytoplasmic RNA from brain, liver and kidney by the methods described by Busch[7] which are commonly used to examine small nuclear RNA species. We separated the RNA samples on acrylamide gels, visualized the RNA by staining with ethidium bromide and, after photography, transferred the RNA to nitrocellulose, and probed the blot with nick-translated p1B224. As indicated in Fig. 2, the commonly observed species (tRNA, 5S RNA, U1–U6, 5.8S and 7S) as well as other previously observed less prevalent RNAs[7] are present in their usual compartments. The ID probe hybridizes to a species which migrates with similar mobility to 5.8S RNA (which is 158 nucleotides) and is located primarily in the brain cytoplasmic fraction. It also hybridizes in a diffuse smear to a target with slightly faster migration than 5S RNA, estimated to be 100–110 nucleotides, which is also found in the cytoplasmic fraction of brain but not liver or kidney.

Our original study showed that the 160-nucleotide RNA was found in the poly(A)$^+$ fraction of brain cytoplasmic RNA prepared by oligo(dT) cellulose chromatography[1]. Other small RNAs have been shown to bind to or co-purify with poly(A)$^+$ RNA[8], so it was necessary to determine whether this small RNA bound directly to oligo(dT) cellulose or indirectly via binding to other poly(A)$^+$ RNA. We denatured total brain cytoplasmic RNA in 90% formamide to disrupt all intermolecular interactions according to the procedure of Jelinek and Leinwand[8], isolated poly(A)-containing RNA by passage over oligo(dT) cellulose, and examined the bound and unbound RNA fractions by Northern blotting using p1B224 as a probe. Most of the small RNA was found in the bound fraction (data not shown), indicating that it binds directly to the oligo(dT) cellulose. Presumably, then, the 160-nucleotide RNA species contain A-rich sequences of sufficient length to permit binding to the oligo(dT)

© 1984 Nature Publishing Group

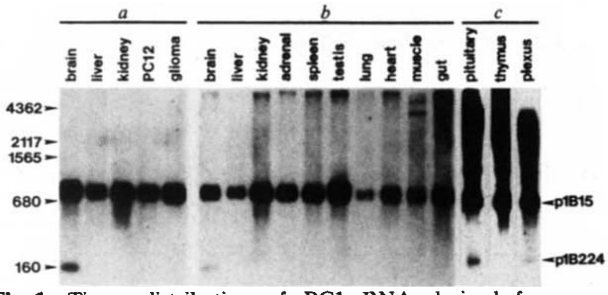


Fig. 1 Tissue distribution of BC1 RNA derived from: *a*, poly(A)$^+$-enriched cytoplasmic RNA (2 μg per lane)[9]; *b*, poly(A)$^+$-enriched total cell RNA (2 μg per lane); *c*, unpurified total cell RNA (20 μg per lane). RNA samples from the indicated rat tissues or cell lines were fractionated by electrophoresis on 1.5% agarose gels in 1M formaldehyde, transferred to nitrocellulose and hybridized with ^{32}P-labelled plasmids p1B224 and p1B15 as described previously[1]. Total cell RNA was prepared as described by Levy *et al.*[21]. The unpurified total cell RNA samples (pituitary, solar plexus and thymus) contain some DNA contamination, accounting for the intense hybridization at high molecular weights. The positions and lengths in nucleotides of DNA size markers (linear pBR322 and *Rsa*I-digested pBR322[13]) are indicated at the left and the positions of RNA bands corresponding to plasmids p1B224 and p1B15 are indicated at the right.

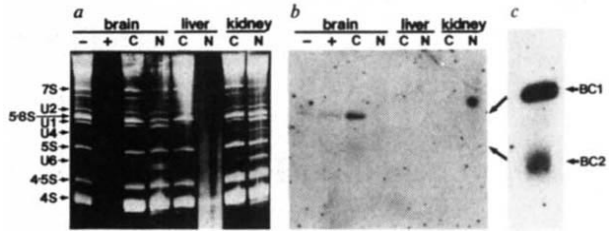


Fig. 2 Localization of BC1 RNA to brain cytoplasm. RNA samples (100 μg per lane, except as indicated) from brain cytoplasmic poly(A)$^-$ (−) or poly(A)$^+$ (+, 10 μg) RNA and total RNA from nuclei (N) or cytoplasm (C) from rat brain, liver or kidney were dissolved in 30% dimethyl sulphoxide, heated to 67°C and separated by electrophoresis on a 10% acrylamide–TBE (50 mM Tris-borate, 1 mM EDTA, *p*H 8.3) gel. The gel was stained with ethidium bromide, photographed (*a*), transferred to nitrocellulose by blotting and the blot hybridized to ^{32}P-labelled p1B224 (*b*). The positions of known small RNA species are indicated in *a* by reference to Reddy *et al.*[7]; *b* is aligned and to the same scale. The smeared pattern of the stained liver nuclear RNA is due either to an excess of salt in the sample or to degradation: similar experiments and the data of Fig. 1 provide no evidence for BC1 RNA in nuclei or cytoplasm of liver. *c* Is an equivalent lane to *b* brain cytoplasm from a separate experiment in which BC2 was more clearly visible. *c* Is shown at a different scale from *b*.

column. The smear in the 100–110-nucleotide range was greatly reduced relative to the 160-nucleotide RNA in oligo(dT)-purified samples.

For convenience, we will call this neural-specific, 160-nucleotide, cytoplasmic, poly(A)$^+$ RNA species BC1 RNA to emphasize its specific location in brain cytoplasm. We will refer to the neural-specific, 100–110-nucleotide species as BC2. When large amounts of brain cytoplasmic RNA purified by oligo(dT) chromatography are examined by gel electrophoresis and ethidium bromide staining, the major species of small RNA observed is a slightly diffuse 160-nucleotide species which migrates just faster than 5.8S RNA. Neither liver nor kidney poly(A)$^+$ RNA exhibits a similar species. From saturation hybridization experiments[9] we estimate that there are about 2,000 copies of BC1 per cell.

ID sequences contain promoters

Two possibilities for the origin of BC1 RNA are that it is a discrete transcript from the genome or that it is derived discretely from brain-specific mRNA processing. The 82-nucleotide ID core contains sequences homologous to the consensus sequence for Pol III initiation[10] (Fig. 3*a*). We studied the transcription of ID plasmid p2A120 in a DNA-dependent Pol III transcription system—a nuclear S100 extract[11] of HeLa cells (Fig. 3*c*). p2A120 was a potent template in this system (10–20 transcripts synthesized per ID sequence per h of reaction), rivalling a cloned *Xenopus* 5S RNA gene (5 transcripts per h) in stimulating transcription. Other ID clones are also excellent templates in this system.

Most of the RNA product from p2A120 ran on the gel in a discrete band of about 250 nucleotides (the 5S product is 120 nucleotides long[12]). A large (~480-nucleotide) p2A120 product, which we will discuss separately, is present at a much lower concentration. We demonstrated that the 250-nucleotide band represents a discrete RNA transcript by cleaving the template plasmid p2A120 before transcription with a variety of restriction endonucleases whose cleavage sites within the p2A120 insert and vector were exactly known from their respective nucleotide sequences[1,13]; these are mapped in Fig. 3*b*. *Pst*I cleavage results in an RNA product of about 110 residues, indicating that Pol III initiation occurs 110 residues from one of the two *Pst*I sites in p2A120. Initiation occurs within the insert because the parent vector pBR322 is not a potent Pol III template (not shown). The amount of synthesis on the *Pst*I-cleaved template and on the *Hpa*II-cleaved plasmid is considerably lower than on the uncleaved plasmid or on the plasmid digested with several enzymes which cut distantly from the *Pst*I site (*Hin*fI, *Sau*3A and *Rsa*I). The *Hpa*II cleavage results in an RNA product of about 165 nucleotides. Enzymes which cleave within the ID sequences, *Bst*N, *Hae*III and *Alu*I (not shown), inactive the template. In some experiments, the inactivation is not complete and some synthesis is observed, often of altered length products. For example, *Ava*II cleavage did not greatly inactivate the template in this or two other experiments, although a shorter product is observed due to cleavage within pBR322 sequences.

All of these data are exclusively consistent with Pol III transcription beginning at or very near the first nucleotide of the ID core sequence (indicated by arrow in Fig. 3*a*), and proceeding for 110 nucleotides to the *Pst*I site and into the pBR322 vector sequences. Cleavage within the ID sequence inactivates this Pol III template, while cleavage just downstream results in a markedly lowered efficiency of initiation. Other bands which show up are attributed to incomplete digestion in some experiments with *Hae*III, and especially *Ava*II. Termination occurs when synthesis runs either to the end of the cleaved template or reaches the Pol III terminator sequence TTTATT (ref. 14) at position 3,472 of pBR322[13], 250 nucleotides downstream from the site of initiation. The 480-nucleotide RNA mentioned above is explained by occasional readthrough of the first terminator to the next T-rich sequence, TTTTT, at vector position 3,242. This readthrough RNA is shortened when the template is cleaved with *Hin*fI or *Sau*3A, as would be expected from the position of these sites in the vector sequence (Fig. 3*b*). Figure 3*b* shows the mapping of the primary and readthrough transcripts. The direction of transcription is the same as our previous determination of the sense strand of the ID sequence for BC1 RNA[1].

Nuclear ID sequence transcription

We next tested nuclei isolated from rat brain, liver and kidney as templates for transcription *in vitro* in the HeLa S100 extract, or in a HeLa whole cell extract[15,16] which contains both Pol II and Pol III activities. The two transcription systems gave essentially analogous results in the following analyses, because the nuclei alone provide both polymerase activities, as evidenced by experiments in which neither extract was provided. Nonetheless, the best incorporation was found using the S100 system. Transcription products containing regions homologous to ID sequences were purified by hybridization to filters containing denatured p2A120 DNA, and the filters were counted (Table 1) and then eluted. The bound transcripts as well as the unbound material were displayed by gel electrophoresis (Fig. 4). Forty-

© 1984 Nature Publishing Group

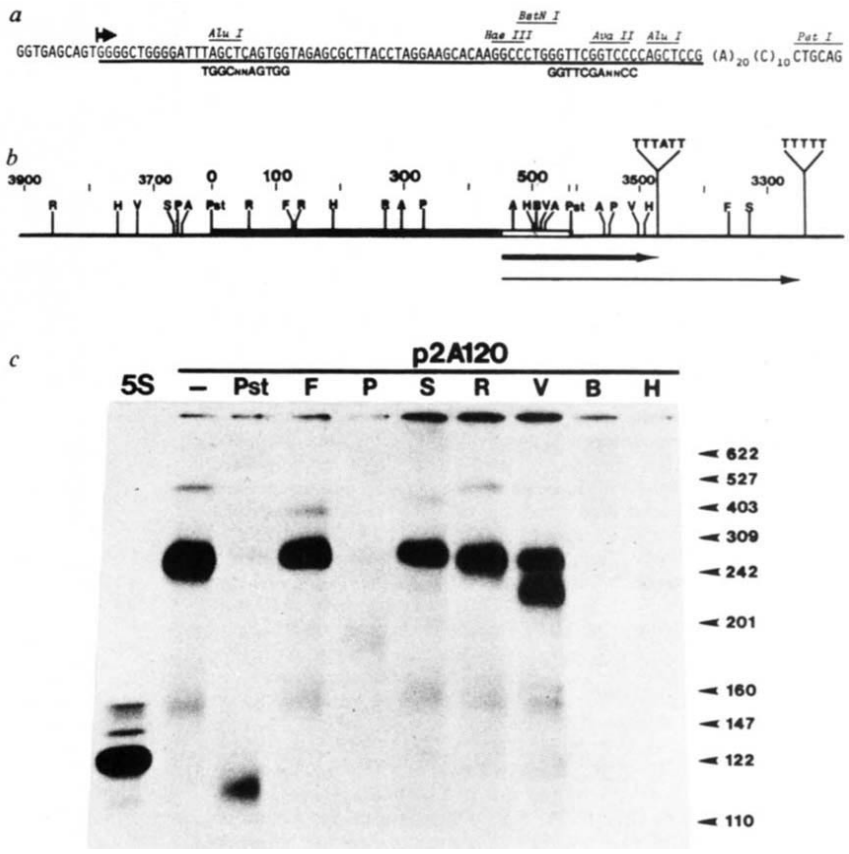


Fig. 3 Transcription of ID plasmid p2A120 *in vitro*. *a*, Nucleotide sequence of the ID region of plasmid p2A120[1]. The core sequence is underlined (some of the non-underlined A residues are part of the core[4]), restriction endonuclease sites are indicated above and the Pol III consensus initiation sequences[10] are indicated below. The site of Pol III initiation is indicated by the arrow. *b*, Structure of plasmid p2A120 showing cDNA insert sequences (thick line) and surrounding pBR322 sequences (thin line). The site of the ID region (corresponding to *a*) is shown by the unfilled thick line. The larger numbers above indicate map positions for insert sequences and the smaller numbers map positions for pBR322 sequences[13]. Letters indicate the sites for restriction enzymes *Pst*I (Pst), *Hin*fI (F), *Hpa*II (P), *Sau*3A (S), *Rsa*I (R), *Ava*II (V), *Bst*NI (B), *Hae*III (H) and *Alu*I (A). Shown below are the direction of transcription and putative positions of the major, 250-nucleotide (thick arrow) and minor, 480-nucleotide (thin arrow) transcripts. The probable termination sequences for these are shown above. *c*, Pol III transcription of plasmid p2A120 *in vitro*. Plasmid DNA was transcribed *in vitro* in a HeLa S100 system[11] as described in Table 1 legend and the products were fractionated by electrophoresis on an 8% acrylamide–7 M urea–TBE gel. The figure shows the products of *Xenopus* clone pX108 (5S) and plasmid p2A120 either untreated (–) or digested with the indicated restriction endonucleases (abbreviated as in *b*). The positions and lengths in nucleotides of DNA size markers are shown on the right (*Hpa*II-digested pBR322[13]).

seven per cent of the counts transcribed from p2A120 DNA rehybridized to p2A120 DNA on filters, thus providing a measure of hybridization efficiency for this experiment. The 250- and 480-nucleotide species discussed above are observed. Nuclei from brain, liver and kidney direct the synthesis of comparable amounts of RNA, much of which runs as a heterogeneous smear all the way to the top of the gel. In addition, a series of discrete bands is observed at 100–110 nucleotides in all three lanes, although these are by far most pronounced in the brain and are not of exactly comparable relative intensities and position in the three nuclear samples. The bands are quite distinct in size from those corresponding to pre-tRNA species synthesized by *Xenopus* oocytes (not shown).

A large fraction (29%) of the brain nuclear transcripts, but hardly any of those derived from liver (2%) or kidney (1%), hybridized to the ID filters. The 100–110-nucleotide discrete brain nuclear transcripts bound to the ID filter, but the species with similar gel mobilities from liver or kidney did not bind at all. Deproteinized rat liver DNA programmes the transcription of a heterogeneous mixture of RNA species of 100–500 nucleotides in length of which 13% hybridize to the ID DNA filter: the DNA did not programme synthesis of the 100–110-nucleotide species. If the liver nuclear extract was treated briefly with 0.35 M NaCl before the reaction, a treatment known to cause the removal and structural rearrangement of some chromosomal proteins[17,18], transcription of large RNA occurred which bound to the ID sequence (Fig. 4, lanes *o–r*). Salt wash did not activate the synthesis of the 100–110-nucleotide RNA species, implying the absence of specific Pol III transcription, as we show later. Therefore, nuclei from all three tissues are active as template in this transcription system, but only brain nuclei or salt-disrupted chromatin directs the synthesis of transcripts containing the ID sequence.

Indeed, a substantial portion of the brain nuclear transcripts (possibly 29%/47% = 62%) must contain an ID sequence, as 29% hybridized to the ID clone, whereas the efficiency of hybridization was only 47% as judged by the hybridization of p2A120 transcripts. Fewer than 4% (a background level) of the liver and kidney transcripts, normalized for hybridization efficiency, contain ID sequences. These measurements might underestimate the per cent of transcripts containing ID sequences, since RNA degradation can separate ID from other brain sequences, thereby reducing hybridization. Several experiments of this sort have been performed in which the absolute percentage of counts bound varies, but normalized values in the range of 60% for brain are usually obtained. The other tissues have always yielded background values. Thus, variability lies in the efficiency of hybridization rather than transcription.

Transcription by polymerases II and III

Even though the above studies indicate that ID sequences are transcribed specifically from ID clones by a Pol III system, we doubt that Pol III can account for the presence of ID sequences in the RNA molecules from which we cloned these cDNA copies. Pol III would be expected to initiate transcription at the ID sequence, whereas our cDNA clones have ID sequences at their 3′ ends. Their appearance in the clones at an end is probably a consequence of priming reverse transcription on the oligo(dA) tract with oligo(dT) in the cDNA clone construction, and hence these ID clones are probably derived from internal portions of larger RNA molecules, most likely Pol II transcripts.

Treatment of the nuclear transcription reactions in S100 or whole cell[15] extracts with α-amanitin (an inhibitor of Pol II at 0.5 μg ml^{-1}) diminished the amount of ID-bound RNA (Fig. 4, lanes *k–n*). It is apparent in several experiments that the synthesis of the large heterogeneous RNA transcripts was inhibited, while that of the smaller (100–110-nucleotide) species was not. Not all transcription of large RNA is inhibited, possibly because of some long Pol III transcription units. These data suggest that most of the large brain-specific transcripts are Pol II products, whereas the smaller brain-specific products are made by Pol III. Thus, both Pol II and Pol III transcripts of brain frequently contain ID sequences.

Because much of the following discussion depends on accepting conclusions drawn from the experiments shown in Fig. 4, and the photograph of the results of lanes *k–r* is unappealing,

© 1984 Nature Publishing Group

Table 1 *In vitro* transcription of ID sequences

Template	Total c.p.m. incorporated	% Bound	Normalized
p2A120 DNA	5,409	46.8	100
Liver DNA	12,593	12.6	26.9
Brain nuclei	3,945	29.4	62.8
Liver nuclei	5,723	2.0	4.1
Kidney nuclei	4,024	1.0	2.2

Nuclei were prepared from brain, liver or kidney of 3-month-old Sprague–Dawley rats by homogenization in 10 volumes of 0.3 M sucrose, 10 mM Tris-HCl *p*H 8, 2.5 mM $MgCl_2$, 5 mM dithiothreitol, 0.25% Triton X100. After removal of cell debris the nuclei were collected by centrifugation at 1,000*g* for 10 min and washed once in homogenization buffer and several times in the transcription buffer. DNA was prepared from liver nuclei by extraction with proteinase K/SDS and phenol/chloroform. Transcription extracts were prepared from HeLa cells by the methods of Weil *et al.*[11]. All extracts were stored at −70°C and thawed once before use. Transcription reactions were carried out in a volume of 50 μl with 20–30 μl of cellular extracts giving final concentrations of 12 mM HEPES (*p*H 7.9), 66 mM KCl, 8 mM $MgCl_2$, 0.06 mM EDTA, 1.2 mM dithiothreitol, 11% glycerol. Template amount per reaction: p2A120, 1 μg; liver DNA, 2.5 μg; nuclei, 20 μg. Incubations were at 22°C for 2 h with 0.02 mM [α-^{32}P]GTP (final specific activity 10 Ci $mmol^{-1}$) and unlabelled nucleotides at 0.6 mM. RNA was purified from the reactions by extraction with phenol/chloroform. Transcripts containing ID sequences were isolated by hybridization to denatured p2A120 bound to nitrocellulose filters (1 μg per filter). Hybridization was in 0.45 M NaCl, 20 mM HEPES *p*H 7.5, 0.5% SDS, 10 mM EDTA, at 55–60°C overnight. Bound material was eluted in 50% formamide at 95 °C for 5 min and concentrated by ethanol precipitation. For each template the table shows the total incorporation into RNA, the % of the total incorporation which bound to p2A120 filters, and the % normalized for hybridization efficiency.

a few comments are required. We have repeated these experiments several times and the qualitative results have always been the same. However, electrophoresis or autoradiographic problems have marred the final product such that photographs of the results require extensive explanation. When we take all of the data together, we are confident of three points: α-amanitin greatly reduces the amount of large (Pol II) transcripts which are synthesized and which bind to ID filters, but has a much smaller inhibitory effect on the smaller (Pol III) transcripts; salt wash activates transcription of ID sequences from liver chromatin, indicating that some chromosomal proteins exert a negative effect on ID expression; and salt wash does not activate the expression from liver of the 100–110-nucleotide RNA species.

Discussion

Our results show that a 160-nucleotide poly(A)$^+$ RNA molecule, BC1, and a smaller 100–110-nucleotide species, BC2, are present in the cytoplasm of brain cells. BC1 is also found in pituitary and peripheral ganglia but not nine other non-neural tissues. Saturation hybridization experiments have shown that there are ~2,000 copies of BC1 per cell[9]. BC1 and BC2 are homologous to brain ID sequences which are specifically transcribed *in vitro* from brain, but not from liver or kidney nuclei, by both RNA polymerases II and III. This specific transcription pattern is not in the main due to DNA rearrangements because naked liver DNA, as well as 0.35 M NaCl-washed liver chromatin, are also templates for the transcription of ID sequences. A cloned ID sequence serves as a template *in vitro* for Pol III transcription and Pol III initiation maps to the first nucleotide of the ID sequence.

Our recent results[4] demonstrate that there is a high correlation of physical linkage in the genome between ID sequences and particular genes which are expressed as brain-specific mRNAs. Brain genes contain ID sequences in their introns. The present study indicates that there is brain-specific transcription of ID sequences, provides further evidence of physical linkage of brain

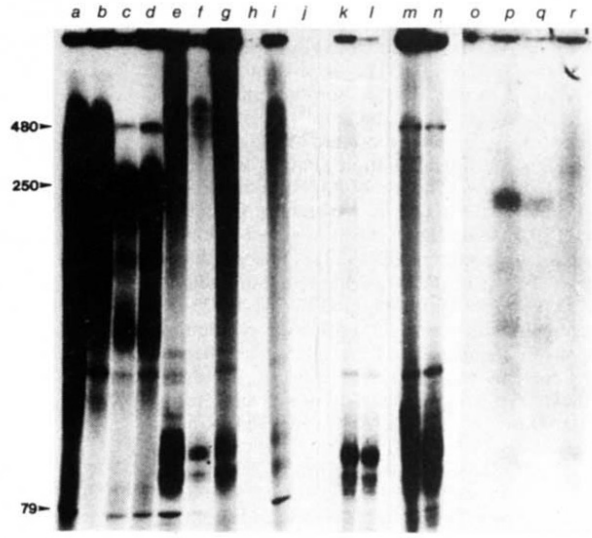


Fig. 4 *In vitro* transcription of ID sequences. RNA was transcribed *in vitro* from DNA or rat nuclei preparations in a HeLa nuclei S100 (Weil)[11] extract and radioactivity labelled by incorporation of [α-^{32}P]dGTP. Transcripts containing ID sequences were isolated by hybridization to filter-bound p2A120 and the products analysed by electrophoresis on an 8% acrylamide–7 M urea–TBE gel. The autoradiographs of the gels for three experiments are shown. In one experiment (lanes *a–j*), the RNA products were transcribed from: deproteinized rat liver DNA (lanes *a, b*); p2A120 (*c, d*), rat brain nuclei (*e, f*), rat liver nuclei (*g, h*), rat kidney nuclei (*i, j*). Lanes *a, c, e, g, i* are the transcription products which did not bind to the p2A120 filters; lanes *b, d, f, h, j* are transcripts which bound to p2A120. Experimental details are described in Table 1 legend. Size markers are indicated at the left. The 79-nucleotide RNA band is a DNA-independent product commonly observed with the S100 extract which binds nonspecifically to nitrocellulose. In a second experiment (lanes *k–n*), brain nuclei were transcribed in the absence (*k, m*) or presence (*l, n*) of 0.5 μg ml^{-1} α-amanitin. The bound (*k, l*) and unbound (*m, n*) transcripts are shown. In a third experiment (lanes *o–r*) liver (*o, q*) or brain (*p, r*) nuclei were transcribed without (*o, p*) or with treatment with 0.35 M NaCl. The bound transcripts are shown. The three autoradiographs have been scaled such that the 100–110-nucleotide species are approximately aligned.

genes and ID sequences, and suggests that ID sequences are largely absent from the introns of genes expressed in non-neural tissues. These data provide circumstantial support for the hypothesis that ID sequences are involved in some event in the generation of neural-specific mRNA, perhaps in the control of transcription.

It is intriguing that ID sequences should be specifically transcribed by both Pol II and Pol III. Other investigators have suggested that transcription by those two polymerases may be somehow coupled: most recently, Manley and Colozzo[19] suggested that Pol III transcription initiating at an *Alu* sequence near a globin gene may be involved in activating Pol II transcription of that globin gene. There are about 10^5 copies of the ID sequence in the rat genome[4] and about 3×10^4 distinct poly(A)$^+$ mRNA molecules expressed in rat brain[9]. Therefore, the present findings that ID sequences are specifically transcribed from brain nuclear extracts *in vitro* by Pol III and Pol II, suggest a model in which Pol III transcription of ID sequences specifically in neurones could somehow activate specific Pol II transcription of the adjacent regions. The observation that salt-washed liver chromatin is activated for brain-specific Pol II but not Pol III transcription suggests the existence of brain Pol III transcription specificity factors which recognize ID sequences. Indeed, preliminary experiments have shown that brain extracts are at least a 10-fold better source of ID-specific Pol III transcription activity than are liver extracts.

Pol III transcription could also account for the existence *in vivo* of the brain-specific RNA molecules, BC1 and BC2. Brain

© 1984Nature Publishing Group

nuclei specifically direct the synthesis by Pol III of a series of discrete, homologous RNAs of 100–110 nucleotides which probably correspond to BC2. These could be early terminating species or perhaps precursors of BC1 which are not polyadenylated in this system. Although there is no precedence for polyadenylation of Pol III transcripts, our evidence suggests that BC1 contains poly(A), and that BC2 does not and is a Pol III product. Post-transcriptional polyadenylation could account for the difference in size of the 110-nucleotide α-amanitin-resistant S100-extract product and the 160-nucleotide BC1 species found *in vivo*. One model for the generation of BC1 and BC2 RNA is that ID sequences with their 3′ oligo(dA) tracts are located in brain genes. These are recognized by neural transcription specificity factors and then transcribed by Pol III, with termination occurring past the oligo(dA) region to generate BC2. Most such BC2 transcripts are unstable. A few of the 10^5 ID sequences, however, may have a T in the oligo(dA) region, forming the sequence AATAAA which can serve as a poly(A) addition signal, thereby producing BC1. The polyadenylated BC1 molecules are transported to the cytoplasm where they form a stable pool of a few thousand molecules which may or may not be functional. Meanwhile, we speculate that the Pol III transcription of ID sequences has activated the genes containing the ID sequence in a necessary manner for Pol II transcription and consequently specific mRNA expression.

Clearly, ID sequences could not be the only factor involved in controlling neural gene expression because the brain is a collection of heterogeneous cell types which are not homogeneous in terms of mRNA expression[9,20]. ID sequences may be necessary but are not sufficient for brain gene expression. They perhaps represent a general primary neural gene signal upon which other control systems operate. For most brain-specific genes, a secondary positive signal is required for expression. Specificity is also mediated by salt-releasable factors which keep transcription by Pol II of regions containing ID sequences at low levels in non-brain tissues, arguing that negative control, acting either at the level of specific repressors or, more likely, at the level of overall chromatin structure, is responsible for keeping neural-specific genetic regions silent in non-neural tissues. ID sequences may be the lineage-specific control elements.

This model, then, utilizes *cis*-acting sequences within introns to identify tissue-specific genes, and *trans*-acting transcription specificity factors to guide Pol III to those genes. A separate switching mechanism must regulate the tissue-specific expression of the transcription specificity factors. Liver and kidney apparently produce transcripts analogous to BC2 but having different sequences. These may be transcripts of the liver and kidney *cis*-acting elements.

Clearly our prejudice in this discussion has been that ID sequences influence transcription; however, they could alternatively or additionally be involved in mRNA processing as nucleation sites for RNA folding or protein–RNA interactions (as discussed by Milner *et al.*[4]). The specificity of BC1 RNA suggests it, too, could carry out a brain-specific role.

We thank Judy Ogata, Chris Clare, Mary Ann Brow and Debbie Andrews for technical contributions. This work was supported in part by grants to J.G.S. from NIH (GM 32355) and McNeil Laboratories, and to J.M.G. from NIH (GM 26453). This is publication no. 2988-IMM of the Research Institute of Scripps Clinic.

Received 30 September; accepted 12 December 1983.

1. Sutcliffe, J. G., Milner, R. J., Bloom, F. E. & Lerner, R. A. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4942–4946 (1982).
2. Barta, A., Richards, R. I., Baxter, J. D. & Shine, J. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4867–4871 (1981).
3. Hojvat, S., Baker, G., Kirsteins, L. & Lawrence, A. M. *Brain Res.* **239**, 543–557 (1982).
4. Milner, R. J., Bloom, F. E., Lai, C., Lerner, R. A. & Sutcliffe, J. G. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
5. Thomas, P. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5201–5205 (1980).
6. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Bery, P. J. *Molec. Biol.* **113**, 237–251 (1977).
7. Reddy, R. *et al. J. biol. Chem.* **256**, 8452–8457 (1981).
8. Jelinek, W. & Leinwand, L. *Cell* **15**, 205–214 (1978).
9. Milner, R. J. & Sutcliffe, J. G. *Nucleic Acids Res.* **11**, 5497–5520 (1983).
10. Galli, G., Hofstetter, H. & Birnstiel, M. L. *Nature* **294**, 626–631 (1981).
11. Weil, P. A., Luse, D. S., Segall, J. & Roeder, R. G. *Cell* **18**, 469–484 (1979).
12. Brown, D. D. & Gurdon, J. B. *Proc. natn. Acad. Sci. U.S.A.* **79**, 2064–2068 (1977).
13. Sutcliffe, J. G. *Cold Spring Harb. Symp. quant. Biol.* **43**, 77–90 (1978).
14. Bogenhagen, D. F. & Brown, D. D. *Cell* **24**, 261–270 (1981).
15. Manley, J. L., Fine, A., Cano, A., Sharp, P. A. & Gefter, M. L. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3855–3859 (1980).
16. Sudgen, B. & Keller, W. *J. biol. Chem.* **248**, 3777–3788 (1973).
17. Spadafora, C., Ondet, P. & Chambon, P. *Eur. J. Biochem.* **100**, 225–235 (1979).
18. Weisbrod, S. & Weintraub, H. *Proc. natn. Acad. Sci. U.S.A.* **76**, 631–635 (1979).
19. Manley, J. L. & Colozzo, M. T. *Nature* **300**, 376–379 (1982).
20. Sutcliffe, J. G., Milner, R. J., Shinnick, T. M. & Bloom, F. E. *Cell* **33**, 671–682 (1983).
21. Levy, D. E., Lerner, R. A. & Wilson, M. C. *Proc. natn. Acad. Sci. U.S.A.* **79**, 5823–5827 (1982).

Point mutations close to the AUG initiator codon affect the efficiency of translation of rat preproinsulin *in vivo*

Marilyn Kozak

Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA

To determine whether sequence context influences the ability of an AUG triplet to be recognized as an initiator codon by eukaryotic ribosomes, single nucleotide changes were introduced near the translational start site in a cloned preproinsulin gene. Maximum synthesis of preproinsulin occurred when a purine, preferably adenosine, was located three nucleotides upstream from the initiator codon. Adenosine is found most frequently in that position in natural mRNAs.

TRANSLATION begins at the 5′-proximal AUG triplet in about 90% of the eukaryotic mRNAs that have been sequenced. That observation, together with other circumstantial evidence reviewed previously[1], prompted the hypothesis that the 40S subunit of eukaryotic ribosomes might find the AUG initiator codon by 'scanning' the 5′ end of messenger RNA. A partial vindication of the scanning model has come from experiments carried out in cell-free extracts. By studying the binding of specially designed templates to wheat germ ribosomes, we have demonstrated that mRNA must have a free 5′ terminus in order to be functional[2], that 40S ribosomal subunits can be trapped upstream from the AUG initiator codon[3] and that 40S subunits can migrate[3,4]. Consistent with a scanning mechanism, the translation of certain viral mRNAs was arrested following hybridization with DNA fragments complementary to a portion of the 5′-noncoding region[5–7]. Recent experiments carried out *in vivo* further support the idea of scanning: ribosomes appear to initiate exclusively at the 5′-proximal AUG triplet in mRNAs that contain tandemly reiterated copies of the preproinsulin ribosome binding site[8]. Although it seems clear that the position of

© 1984Nature Publishing Group

an AUG triplet relative to the 5′ end of the mRNA has a major role in identifying it as a functional initiator codon, the notion of scanning is not incompatible with the idea that sequence context modulates the efficiency with which an AUG triplet is recognized as an initiator codon. Indeed, the distribution of nucleotides flanking functional initiator codons in eukaryotic mRNAs is decidedly nonrandom. From a comparison of the 5′-noncoding sequences of over 200 cellular[9] and some 50 viral mRNAs[10], $CC^{A}_{G}CCAUGG$ has recently been tentatively proposed as a consensus sequence for eukaryotic initiation sites. The most highly conserved features revealed by those surveys were adenosine (A) in position −3, (that is, three nucleotides upstream from the AUG triplet, which is numbered +1 to +3) and guanosine (G) in position +4. The significance of the conserved nucleotides flanking the AUG triplet has been tested to a limited extent by measuring the binding of synthetic oligonucleotides to ribosomes *in vitro*[10]. To assess further whether the translational machinery in eukaryotes is sensitive to the sequence context around a potential AUG initiator codon, I have introduced single nucleotide changes near the translational start site in a cloned preproinsulin gene. The yield of proinsulin was then monitored during transient expression of the mutant plasmids in monkey (COS-1) cells. Sequence changes around the initiator codon drastically affected the yield of proinsulin, as shown in the experiments described below.

Plasmid construction

Plasmid 255 (p255) is a shuttle vector that encodes and expresses the rat preproinsulin II gene. (The plasmid was donated by Dr Peter Lomedico.) Its structure has been described previously[11] and is briefly outlined in Fig. 1*a*. Initiation of transcription at the simian virus 40 (SV40) early promoter in p255 gives rise to a chimaeric mRNA with a 5′-untranslated sequence of about 128 nucleotides: the first ~80 nucleotides are encoded by SV40 DNA and the remainder are from the rat insulin gene. A *Hin*dIII site marks the boundary between SV40 and rat insulin DNA, as shown in Fig. 1*a*, *b*. Rather than attempt to introduce mutations around the natural preproinsulin initiator codon, which lacks nearby restriction sites, I inserted upstream from the insulin coding sequence an 11-base pair (bp) oligonucleotide that carries another ATG triplet. The resulting mutant (p255/2) is illustrated in the second line of Fig. 1*c*, with the 11-nucleotide insert shown in bold face. To facilitate the construction and identification of additional mutants, the synthetic oligonucleotide was designed with *Hin*dIII and *Bcl*I restriction sites flanking the ATG triplet. Because the upstream ATG triplet in p255/2 lies in the same reading frame as the preproinsulin coding sequence, I anticipated that, after introducing mutations around the upstream ATG codon, I would be able to monitor initiation at the first versus the second ATG by measuring the size of the insulin-related polypeptides: the polypeptide resulting from initiation at the upstream ATG should carry an extra 18 amino acids at the N-terminus. That experimental design proved impractical, however, because the long form of preproinsulin encoded by p255/2 was still an efficient substrate for signal peptidase (M.K., unpublished data). Thus, irrespective of which ATG was used for initiation, the only polypeptide that accumulated was the cleavage product, proinsulin.

To obviate that problem, another derivative (p255A/1) was constructed, as illustrated in the bottom line of Fig. 1*c*. p255/2 DNA was cut with *Bcl*I and *Bam*HI which generate identical cohesive ends (GATC), although each enzyme recognizes a different hexameric sequence. The large DNA fragment was then re-ligated, that is, the *Bcl*I-generated end was joined to the *Bam*HI-generated end, thereby deleting 182 bp including the ATG triplet that is normally used to initiate translation of preproinsulin. In place of the natural initiator codon, p255A/1 has the ATG triplet derived from the oligonucleotide insert in p255/2. The sequence around the new initiator codon in p255A/1 is shown in Fig. 1*c*. p255A/1 directs synthesis of a polypeptide which lacks only three amino acids that are normally present at the N-terminus of preproinsulin. In the remainder of this article, p255A/1 is used as the point of reference: a series of mutants differing from p255A/1 by single nucleotide changes is described below, and their translational properties are compared.

Introduction of mutations

Site-specific mutagenesis was carried out by first forming gapped heteroduplex circles in which the region around the initiator codon was single-stranded (see Fig. 2*a*, *b*), and then annealing to the gap a complementary oligonucleotide that contains a single mismatched residue. The mismatched nucleotides are starred in Fig. 2. Two mutants were obtained using scheme *a*: p255/15 and p255/16 are identical to p255/2 (described above) except that p255/15 has an A and p255/16 has a G, in the −3 position. The two new mutants could not be used directly in the protein synthesis assay because the polypeptide resulting from initiation at the upstream ATG triplet would undergo cleavage, a problem explained in the preceding discussion of p255/2. Therefore, the technique whereby p255A/1 was derived from p255/2 was applied again: p255/15 was cut with *Bcl*I and *Bam*HI and then re-ligated, generating p255A/3; p255/16 was cut similarly and re-ligated to produce p255A/4. Derivatives p255A/3 and p255A/4 are identical in structure to p255A/1, except for the nucleotide in the −3 position. The sequence around the initiator codon in each plasmid is given in Table 1.

A slightly different scheme was needed to obtain mutants in the +4 position, because changing the A in that position would destroy the *Bcl*I site and thus preclude the *Bcl*I/*Bam*HI fusion step used to derive the aforementioned mutants. Derivative p255A/2, in which the sequence ATGA has been changed to ATGG, was obtained using scheme *b* in Fig. 2. The technique again involved using a mismatched oligonucleotide to complement the single-stranded region in a gapped heteroduplex circle, but the parental plasmids used to construct the heteroduplex in scheme *b* differ from those in scheme *a*. Finally, scheme *c* in Fig. 2 illustrates the approach used to obtain a double mutant (p255A/7) in which the nucleotides in both the −3 and +4 positions have been changed to G. The technique here simply involved forming a nicked heteroduplex circle between two of the mutants already in hand, p255A/4 and p255A/2. The mismatched nucleotide in the −3 position of the heteroduplex underwent correction *in vivo* by the bacterial DNA repair machinery, generating the desired double mutant.

All of the mutants resulting from the constructions outlined in Fig. 2 were initially identified by the loss of the *Hin*dIII restriction site that had been present in the parental plasmids, or by the gain of an *Ava*II restriction site ($GG^{A}_{T}CC$) immediately following the ATG codon. The precise structure of each mutant was subsequently confirmed by DNA sequence analysis. The relevant portions of the sequencing gels are shown in Fig. 3.

Expression of mutant plasmids

Synthesis of proinsulin was monitored following transfection of COS-1 cells[12] by p255A/1 or one of the derivatives described above. Transfection of subconfluent cell monolayers was carried out using the standard $CaPO_4$–DNA co-precipitation technique of Wigler *et al.*[13], as described previously[8]. Two days after transfection the cells were labelled with ^{35}S-cysteine, lysed, and the insulin-related polypeptides recovered by precipitating with anti-insulin antibodies, all according to published procedures[8]. The product that accumulates when a cloned preproinsulin gene is introduced into cultured monkey cells has been reported by other workers to be proinsulin[14–16]. In the present study, identification of the proinsulin band rests on four criteria: it is absent from cells transfected with a control plasmid, p255/0, which lacks the entire insulin coding sequence; it is precipitated by anti-insulin antibodies but not by preimmune serum (not shown); it is competed by including excess nonradioactive insulin in the immunoprecipitation reaction (Fig. 4, lane 8); and its apparent molecular weight of ~9,000 corresponds approximately to that of proinsulin.

© 1984Nature Publishing Group

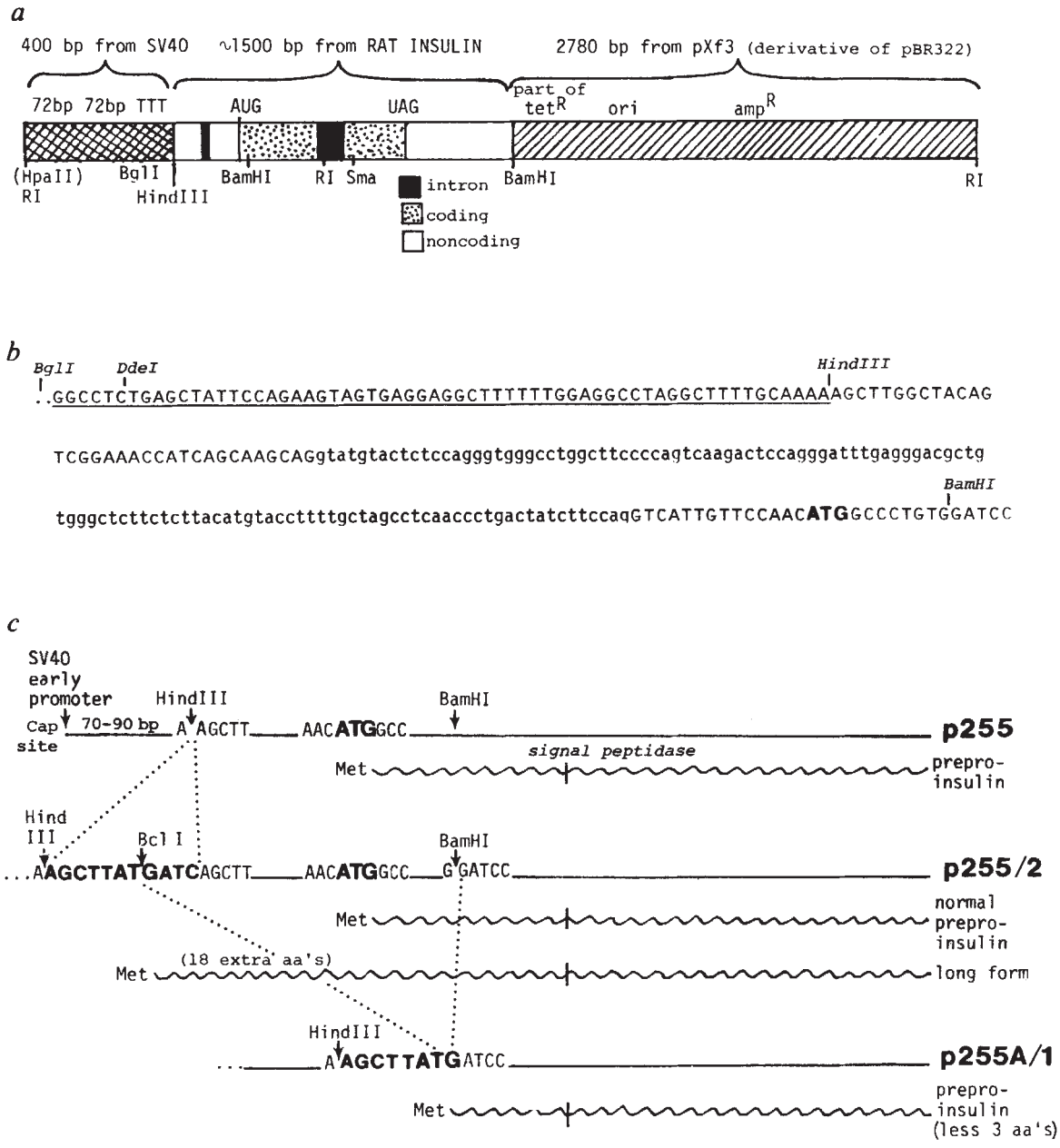


Fig. 1 Structure of transcripts and insulin-related polypeptides encoded by p255 and its derivatives. ***a***, Linear representation of the parental plasmid p255. The segment of SV40 DNA that is included in the plasmid extends from the *Hpa*II site at position 346 to the *Hin*dIII site at position 5,171, 8 bp upstream from the ATG initiator codon of T antigen. The plasmid contains a single *Hin*dIII site, which is at the junction of SV40 and rat insulin sequences. Some of the constructions described in the text were facilitated by eliminating the *Bam*HI site at the junction between rat genomic and pBR322 sequences in p255. This was done by obtaining unit-length linear DNA molecules after partial digestion with *Bam*HI, filling in the sticky ends using *Escherichia coli* DNA polymerase, and then re-ligating. The resulting plasmid (designated p255/11) has a single *Bam*HI site, 8 bp downstream from the ATG initiator codon of preproinsulin. In all other respects p255/11 is identical to wild-type p255. ***b***, Nucleotide sequence preceding and extending into the insulin coding region in p255. The DNA strand that has the same sense as mRNA is shown. The transcript produced by p255 initiates near the *Bgl*I site of SV40 and carries at its 5′ end 70–90 nucleotides encoded by SV40 DNA (underlined). The remainder of the transcript is derived from the rat insulin gene. The portion of the rat insulin sequence that is retained in mature mRNA is printed in capital letters without underlining; the 119-bp intron that interrupts the 5′-noncoding region is shown in lower case letters. The ATG codon that normally initiates translation of rat preproinsulin is shown in bold face. ***c***, Genesis of plasmids p255/2 and p255A/1. The upper line represents a portion of the parental plasmid p255. The *Hin*dIII and *Bam*HI sites are those identified in ***b***. The second entry depicts p255/2, which was derived by inserting an 11-bp oligonucleotide (shown in bold face) at the unique *Hin*dIII site in p255. The third entry shows the structure of p255A/1, in which the sequence between the *Bcl*I and *Bam*HI sites has been deleted. The polypeptides encoded by each plasmid are represented by wavy lines.

Methods: The oligonucleotides 5′-AGCTTATGATC-3′ and 5′-AGCTGATCATA-3′ were custom-synthesized by PL Biochemicals, Inc. When annealed, they form a partially base-paired structure with 5′-protruding termini that are complementary to the overhangs generated by cutting DNA with *Hin*dIII. Derivative p255/2 was obtained by digesting the parental plasmid p255 with *Hin*dIII and alkaline phosphatase, followed by ligation with the aforementioned oligonucleotides which were present in approximately 50-fold molar excess. Plasmids that had acquired the 11-bp insert were detected initially by screening for the *Bcl*I cleavage site (TGATCA). Derivatives with the insert oriented correctly (that is, as illustrated in the figure) were subsequently identified by DNA sequence analysis. The deletion derivative p255A/1 was obtained by sequentially digesting p255/2 DNA (grown in the *dam*$^-$ host *E. coli* SR19) with *Bcl*I and *Bam*HI, and then re-ligating the linear DNA molecules. Plasmids from which the 182-bp *Bcl*I/*Bam*HI fragment had been deleted were initially identified by loss of the *Bcl*I and *Bam*HI restriction sites. The structure shown for p255A/1 was subsequently confirmed by DNA sequence analysis.

© 1984Nature Publishing Group

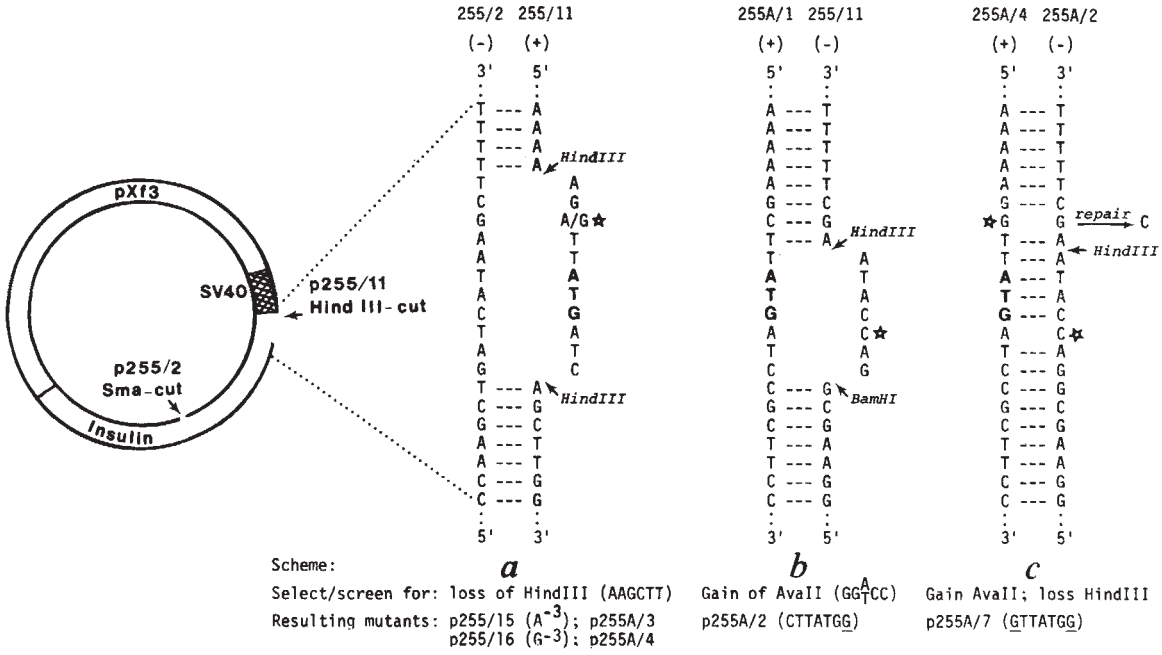


Fig. 2 Construction of point mutants. In protocols *a* and *b*, gapped heteroduplex circles were formed between two plasmid DNA strands of slightly dissimilar lengths, analogous to the procedure described in refs 19–21. The gapped circles were then annealed with an oligonucleotide that was complementary to the single-stranded region except for a single mismatched nucleotide, which is starred. In scheme *a*, for example, *Hin*dIII-cut p255/11 (which is identical to wild-type p255 except for loss of the *Bam*HI site at the junction between rat genomic and pXf3 DNA) and *Sma*I-cut p255/2 DNA were mixed, denatured and then annealed. The resulting heteroduplex circles, illustrated at the far left of the figure, have a single-stranded gap at the position of the 11-bp insertion in p255/2. In schemes *b* and *c*, the DNA strand shown on the left side of the heteroduplexes was linearized with *Sma*I, as in scheme *a*, but different nucleases (identified in the figure) were used to linearize the strand on the right. Recovery of the parental plasmids p255A/1 and p255A/4 was minimized in schemes *b* and *c* by using DNA from plasmids grown in the *dam*$^-$ host, *E. coli* SR19 (ref. 22). As shown in scheme *c*, the double mutant p255A/7 was obtained by forming a nicked heteroduplex circle between two of the mutants already in hand. *E. coli* cells were transformed with the nicked circles and repair occurred *in vivo* at the site of the mismatched nucleotide in position −3, just beyond the *Hin*dIII cut in the DNA strand derived from p255A/2. Each of the illustrated schemes gave only a low yield of the desired mutant; most of the ampicillin-resistant colonies were derived from one or the other unmodified parental plasmid. A portion of the parental-type background is attributable to the formation of gapped heteroduplex circles complementary to those illustrated in the figure. In scheme *a*, for example, the plus-strand of p255/2 and the minus-strand of p255/11 would form a heteroduplex, similar in structure to the one illustrated for the minus-strand of p255/2 and the plus-strand of p255/11. For reasons of fiscal economy, I did not include in the reaction the oligonucleotide required to complement and mutagenize the gapped sequence in the 'complementary' circle, which therefore regenerated parental-type DNA. Despite the high background of parental-type plasmids, the desired mutants were readily identified by the loss of the *Hin*dIII restriction site preceding the ATG codon, or by the gain of an *Ava*II cleavage site following the ATG triplet. The protocol outlined in scheme *a* gave rise directly to mutants designated p255/15 and p255/16, which are identical to p255/2 (described in Fig. 1) except for the nucleotide in the −3 position. From p255/15 and p255/16, derivatives p255A/3 and p255A/4 were subsequently obtained by fusing the *Bcl*I site to the *Bam*HI site, as illustrated for p255A/1 in Fig. 1*c*. **Methods:** Linear DNA from two plasmids (identified in the diagrams), each cleaved with a different single-cutting restriction enzyme, was denatured by boiling for 1 min and then annealed (at 50 μg ml^{-1}) overnight in 20 mM Tris-HCl (*p*H 7.7), 150 mM NaCl and 0.2 mM EDTA, in a water bath that was gradually cooled from 85 °C to room temperature. The reannealed DNA was concentrated by ethanol precipitation and redissolved in a small volume of buffer. 1 μg of the DNA was mixed with 0.02 A_{260} units of the 5′-phosphorylated oligonucleotide that is shown, in the diagram, slightly to the right of the heteroduplex. The oligonucleotides used in schemes *a* and *b* were custom-synthesized by PL Biochemicals, Inc. The DNA/oligonucleotide mixture in 100 μl of 50 mM Tris-HCl (*p*H 7.6), 10 mM $MgCl_2$ and 150 mM NaCl was incubated at 65 °C for 5 min and then held at 15 °C for 16 h. T_4 DNA ligase (4×10^3 units ml^{-1}, New England BioLabs) was present during the 15 °C incubation, together with 1 mM ATP, 20 mM dithiothreitol and bovine serum albumin at 50 μg ml^{-1}. The ligase reaction was stopped by heating at 65 °C for 10 min. An aliquot of the reaction mixture was used directly to transform *E. coli* MM294, according to standard procedures[23]. Ampicillin-resistant colonies were picked and the DNA from 10-ml minicultures[24] was screened with *Hin*dIII or *Ava*II to identify the desired mutants.

The yield of proinsulin varied with each of the mutants tested in Fig. 4. Comparison of lanes 1–3 shows that p255A/3 synthesized ~15-fold more proinsulin than did p255A/1; the translational efficiency of p255A/4 fell between those two extremes. To control for possible variations in transfection efficiency from plate to plate, each cell monolayer was transfected in this experiment with a mixture of two plasmids: the mutant plasmid under investigation (for example, p255A/1) and a previously characterized derivative called p255/10. The salient feature of p255/10 is that it directs synthesis of an elongated form of preproinsulin, as shown in lane 6 of Fig. 4. Using the high molecular weight insulin-related polypeptide encoded by p255/10 as an internal control, the amount of proinsulin produced by p255A/1 and each of its derivatives can be reliably compared. The data in Fig. 4 indicate that the mutant plasmids direct synthesis of proinsulin with vastly different efficiencies. Table 1 summarizes the sequences and the relative translational efficiencies of the various plasmids in the A-series.

The targeted mutagenesis technique used in this study is unlikely to produce secondary mutations outside the region of interest. Although the oligonucleotides used for mutagenesis are rather short, the structure of the gapped heteroduplex circle precludes binding of the oligonucleotides to extraneous sites. Nevertheless, it seemed desirable to prove that the observed variations in translation were due solely to sequence differences around the ATG initiator codon. Towards that end, I subjected the low-yielding plasmid p255A/2 to a further mutagenesis step, cleaving the DNA first with *Hin*dIII and then using S_1

© 1984 Nature Publishing Group

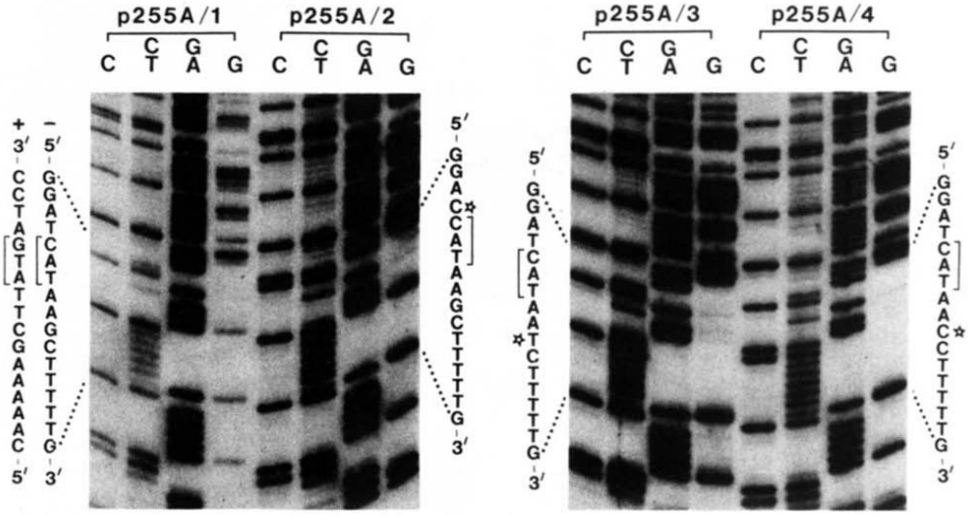


Fig. 3 Sequence analysis of mutant plasmids. The gels show the sequence of the minus-strand of a portion of the DNA near the start site for translation. The first mutant in this series, p255A/1, is shown at the far left; for convenience, a portion of the plus-strand sequence of p255A/1 is written beside the minus-strand. In the sequences of the other mutants, the nucleotides that differ from p255A/1 are starred. The ATG triplet or its complement is bracketed. **Methods:** The 250-bp fragment used for sequencing extends from the *Dde*I site in SV40 DNA (56 bp upstream from the unique *Hin*dIII site shown in Fig. 1*b*) to an *Eco*RI site in the large intron of the rat insulin gene. After cutting the plasmid DNA with *Dde*I, the indented 3′ termini were labelled by incubating with TTP, [α-^{32}P]dCTP (800 Ci mmol^{-1}; NEN) and *E. coli* DNA polymerase I (Klenow fragment, New England BioLabs). The fragment of interest was recut with *Eco*RI and purified by electrophoresis through a 5% polyacrylamide gel. Sequencing was carried out by the chemical cleavage method of Maxam and Gilbert[25] using 8% polyacrylamide gels. The gels were dried before autoradiography.

nuclease to delete five nucleotides directly in front of the ATG codon; the new derivative is called p255A/6. The yield of proinsulin from p255A/6 was at least 15-fold greater than from p255A/2. Thus, inefficient expression of proinsulin by p255A/2 can be attributed to the sequence context around the initiator codon rather than to a spurious second mutation elsewhere in the cloned gene.

Discussion

The experiments presented above provide the first evidence that sequence context influences the ability of an AUG triplet to be recognized as an initiator codon by eukaryotic ribosomes. p255A/1 and p255A/3 provide the most dramatic contrast: changing the nucleotide in the −3 position from C to A enhanced translation of rat proinsulin at least 15-fold. The approximately three-fold difference in translation observed between p255A/3 and p255A/4 indicates that A functions better than G in the −3 position, although either purine in that position is far superior to cytidine. The contribution of the nucleotide in the +4 position is less clear, since p255A/4 and p255A/7 differed by at most 1.5-fold in their production of proinsulin. Attempts are underway to replace the purine in the +4 position with a pyrimidine, which might have more dramatic consequences.

Although I have tested so far only the −3 and +4 positions, which are the most highly conserved positions in natural mRNAs[10], it seems likely that other nearby nucleotides also contribute. I have detected measurable differences in translational efficiency among a number of synthetic initiation sites, all of which have A in the −3 position but differ in the adjacent nucleotides (M. K., unpublished data). Obviously a more extensive set of systematically derived point mutants is needed to determine the full extent of the flanking region that contributes to a 'favourable' initiation site. It also seems likely that structural features in mRNA apart from the environs of the AUG triplet can influence translational efficiency. Thus, the prediction that emerges from the present study must be carefully circumscribed: all other things being equal, translational efficiency should increase dramatically when a purine replaces a pyrimidine in the −3 position; and A in that position works considerably better than G. However, the experiments described here do not predict that any two messages with identical sequences flanking the initiator codon will necessarily be translated with identical efficiencies, since some other feature (such as the sequence adjacent to the cap, or the secondary structure of the message, or the pattern of codon usage) might differ in a way that affects messenger function. Defining those ancillary features is a task for the future.

Additional experiments are also needed to determine precisely what happens when the migrating 40S ribosomal subunit encounters an AUG triplet in an unfavourable sequence context—which, for the time being, can be taken to mean a C in the −3 position. One possibility is that most of the 40S subunits detach from the message. Abortive initiation with release of 40S subunits would account for the observed low yield of proinsulin from p255A/1 and p255A/2. An alternative mechanism is that 40S subunits remain bound to the message and simply bypass an AUG triplet that occurs in an unfavourable context. To evaluate these two possibilities (which are not mutually exclusive), I have constructed plasmids with an ATG triplet upstream and in a different reading frame from the preproinsulin start site, and have monitored the yield of proinsulin as a function of varying the sequence around the upstream ATG triplet. Those experiments are described elsewhere[17].

Inspection of the data in Fig. 4 reveals that p255A/3, which is the best of the mutants analysed in this series, directs synthesis

Table 1 Relative translational efficiency of A-series mutants

Plasmid	Sequence at start of preproinsulin coding region	Relative yield of proinsulin
p255A/1	CAAAAAGCTTATGATCCG (−3, +4)	1
p255A/4	CAAAAAG*GTTATGATCCG	5
p255A/3	CAAAAAG*ATTATGATCCG	15
p255A/2	CAAAAAGCTTATG*GTCCG	1–1.5
p255A/7	CAAAAAG*GTTATG*GTCCG	5–7.5

The underlined sequence TGATCC in p255A/1 is the site of the *Bcl*I–*Bam*HI fusion described in Fig. 1 legend. The *Hin*dIII recognition site (AAGCTT) occurs just upstream from the ATG triplet in p255A/1. In the sequences shown for the other mutants, the nucleotides that differ from p255A/1 are starred. The yield of proinsulin from p255A/1 has been set arbitrarily at 1. The relative yield of proinsulin was determined by running, on a single gel, serial dilutions of each sample. All samples could thus be quantitated from a single autoradiogram. This gave more reproducible results than I was able to obtain by applying the samples (undiluted) to a gel and then comparing a series of autoradiograms exposed for different lengths of time.

© 1984 Nature Publishing Group

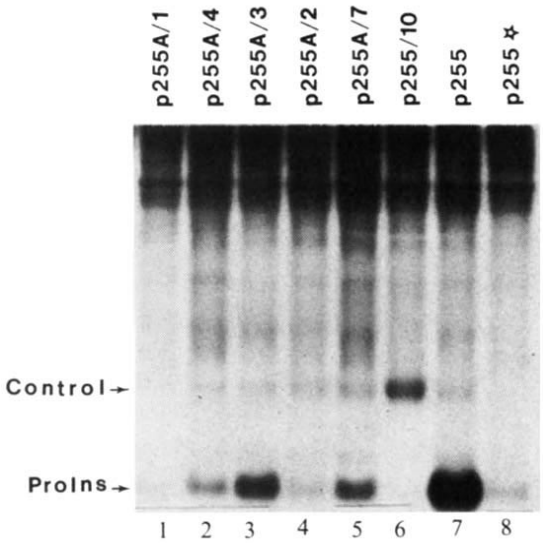


Fig. 4 Variations in proinsulin synthesis as a function of sequence context around the AUG initiator codon. Lanes 1–5 show the yield of proinsulin from cells transfected by mutants in the A-series. For comparison, proinsulin synthesis in cells transfected by wild-type p255 is also shown (lane 7). The proinsulin band is marked 'ProIns', and the high molecular weight insulin-related polypeptide encoded by p255/10 (see text) is marked 'Control'. The intensity of the control band is lower in other lanes, relative to lane 6, due to competition by the test plasmids. The important point, however, is that the intensity of the control band is constant in lanes 1–5 (this was confirmed after prolonged exposure of the autoradiogram), whereas the yield of normal-sized proinsulin varies. For the control shown in lane 8, excess unlabelled insulin was included during the incubation with antiserum, preventing precipitation of both proinsulin and the high molecular weight band encoded by p255/10.

Methods: COS-1 cells were transfected[8,13] with a mixture of the control plasmid p255/10 (5 μg per plate) and 20 μg of one of the test plasmids indicated across the top of the figure. The sample in lane 6 is from cells transfected only with p255/10. Forty-eight h after DNA uptake, the cells were labelled for 3.5 h with ^{35}S-cysteine (NEN, 1,000 Ci mmol^{-1}). Subsequent extraction and immunoprecipitation of the labelled proteins were carried out as described previously[8]. The polypeptides were fractionated on a 15% polyacrylamide-SDS gel[26] containing 6 M urea.

of significantly less proinsulin than does wild-type p255. Several factors might account for this: (1) The polypeptide encoded by p255A/3 and the other plasmids in the A-series lacks three amino acids that are present at the amino-terminus of the natural form of rat preproinsulin. This difference in structure might affect the intracellular processing or stability of the polypeptide. The difference in proinsulin yield between wild type and p255A/3 persists, however, even when the proteins are analysed after a very short pulse with ^{35}S-cysteine. (2) The production and/or stability of mature mRNA might be lower in the case of p255A/3 than in the case of wild-type p255. This explanation seems unlikely, since Lomedico and McAndrew[11] detected no transcriptional anomalies when they perturbed the 5′ structure of the rat preproinsulin gene. (3) Although both the wild-type preproinsulin gene carried by p255 and the mutant form carried by p255A/3 have an A residue three nucleotides upstream from the ATG initiator codon, the sequences around the initiator codons in the two plasmids are otherwise quite dissimilar: the wild-type sequence is CCAACATGG, versus AGATTATGA for p255A/3. I suspect that this difference underlies the more efficient expression of the wild-type preproinsulin gene; but that suspicion awaits verification. It is important to recognize that the conclusions drawn from this study derive not from comparing the wild-type gene with any of the mutants, but from comparing one mutant with another within the A series.

The finding that nucleotides flanking the AUG initiator codon modulate translational efficiency is not unexpected. Although the scanning model has always emphasized the importance of position (that is, proximity to the 5′ end of the message) in defining the functional initiation site, the model was recently revised in the light of two observations. (1) Although translation usually begins at the 5′-proximal AUG triplet in eukaryotic mRNAs, there is a growing list of exceptional cases (~5% of cellular mRNAs and ~10% of viral mRNAs) in which one or more AUG triplets occur some distance upstream from the start of the protein coding sequence[9,10,18]. Thus, eukaryotic ribosomes do not always initiate at the first AUG triplet in the message. (2) The distribution of nucleotides flanking functional initiator codons in eukaryotic mRNAs shows a striking bias in position −3 (where A occurs in ~80% of the messages examined) and in position +4 (where G occurs in 40–60%). The nucleotide preference in positions −1, −2, −4 and −5 is less striking, but a predominance of C has been noted in those positions[9,10]. The nonrandom distribution of nucleotides suggests that the initiation machinery is not indifferent to the sequence context around the AUG triplet. Moreover, the nucleotides—most notably, a purine in the −3 position—that distinguish functional initiator codons rarely occur around the 'nonfunctional' upstream AUG triplets in the exceptional mRNAs cited above. These observations were incorporated into a modified scanning model that emphasizes the importance of both position and sequence context in identifying the initiation site. The modified scanning model[10,18] states that a 40S ribosomal subunit binds initially at the 5′ end of mRNA and subsequently migrates until it reaches the first AUG triplet: if the first AUG triplet occurs in an optimal sequence context, all 40S subunits stop there and that AUG serves as the unique initiator codon, but if the first AUG triplet occurs in a suboptimal sequence context, only some 40S subunits stop and initiate there; some bypass that site and initiate at another AUG that lies farther downstream. The experiments described here represent the first step towards defining the 'optimal sequence context' for initiation by eukaryotic ribosomes.

I thank Peter Lomedico for providing the parental plasmid p255. This research was supported by a grant from the NIH. M.K. is the recipient of Research Career Development Award AI00380.

Received 22 September; accepted 1 December 1983.

1. Kozak, M. *Cell* **15,** 1109–1123 (1978); **22,** 7–8 (1980); *Curr. Topics Microbiol. Immun.* **93,** 81–123 (1981).
2. Kozak, M. *Nature* **280,** 82–85 (1979).
3. Kozak, M. *Cell* **22,** 459–467 (1980).
4. Kozak, M. & Shatkin, A. J. *J. biol. Chem.* **253,** 6568–6577 (1978).
5. Privalsky, M. L. & Bishop, J. M. *Proc. natn. Acad. Sci. U.S.A.* **79,** 3958–3962 (1982).
6. Perdue, M. L., Borchelt, D. & Resnick, R. J. *J. biol. Chem.* **257,** 6551–6555 (1982).
7. Preston, C. M. & McGeoch, D. J. *J. Virol.* **38,** 593–605 (1981).
8. Kozak, M. *Cell* **34,** 971–978 (1983).
9. Kozak, M. *Nucleic Acids Res.* (in the press).
10. Kozak, M. *Nucleic Acids Res.* **9,** 5233–5252 (1981).
11. Lomedico, P. T. & McAndrew, S. J. *Nature* **299,** 221–226 (1982).
12. Gluzman, Y. *Cell* **23,** 175–182 (1981).
13. Wigler, M., Pellicer, A., Silverstein, S. & Axel, R. *Cell* **14,** 725–731 (1978).
14. Lomedico, P. T. *Proc. natn. Acad. Sci. U.S.A.* **79,** 5798–5802 (1982).
15. Gruss, P. & Khoury, G. *Proc. natn. Acad. Sci. U.S.A.* **78,** 133–137 (1981).
16. Laub, O. & Rutter, W. J. *J. biol. Chem.* **258,** 6043–6050 (1983).
17. Kozak, M. *Nucleic Acids Res.* (submitted).
18. Kozak, M. *Microbiol. Rev.* **47,** 1–45 (1983).
19. Kalderon, D., Oostra, B. A., Ely, B. K. & Smith, A. E. *Nucleic Acids Res.* **10,** 5161–5171 (1982).
20. Oka, A., Sugimoto, K., Sasaki, H. & Takanami, M. *Gene* **19,** 59–69 (1982).
21. Peden, K. W. C. & Nathans, D. *Proc. natn. Acad. Sci. U.S.A.* **79,** 7214–7217 (1982).
22. Kramer, W., Schughart, K. & Fritz, H-J. *Nucleic Acids Res.* **10,** 6475–6485 (1982).
23. Morrison, D. A. *Meth. Enzym.* **68,** 326–331 (1979).
24. Klein, R. D., Selsing, E. & Wells, R. D. *Plasmid* **3,** 88–91 (1980).
25. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65,** 499–560 (1980).
26. Laemmli, U. K. *Nature* **227,** 680–685 (1970).

© 1984 Nature Publishing Group

Orientation of nucleosomes and linker DNA in calf thymus chromatin determined by photochemical dichroism

Sekhar Mitra, Dipankar Sen & Donald M. Crothers

Departments of Chemistry and Molecular Biophysics & Biochemistry, Yale University, PO Box 6666, New Haven, Connecticutt 06511, USA

The dichroism for photochemical attachment of a psoralen derivative to Mg^{2+}-stabilized chromatin fibres is used to deduce the orientation of nucleosomal disks and linker DNA in the 30-nm fibre. The new technique of photochemical electric dichroism should have general applicability to problems of nucleic acid organization in cellular subunits and viruses.

DNA in eukaryotic nuclei is packaged by folding in a hierarchical series of progressively more compact structures, ranging from nucleosomes to the metaphase chromosome. The fundamental nucleosomal subunit of chromatin contains 166 base pairs (bp) of DNA negatively supercoiled around an octamer of histones, and has associated with it a single molecule of the lysine-rich histone H1 (ref. 1). Considerable progress has been made in studies of the three-dimensional structure of the nucleosomal core particle[2–4], which is obtained from the nucleosome upon removal of histone H1 and further nuclease digestion to a DNA size of 146 bp.

Polynucleosomes in low salt concentration appear as a fibre about 10 nm in diameter. Upon extension, an alternating arrangement of 166 bp nucleosomes and linker DNA segments becomes evident[5]. The linkers are of nonuniform length, up to 80 bp in size depending on the source of the chromatin sample[6].

Monovalent ion concentrations exceeding about 50 mM[7], or divalent ion concentrations above 0.1 mM[8,9], cause the 10 nm fibre to fold into a more compact conformation, producing a fibre of about 30 nm diameter. Histone H1 is required for this condensation[7,9]. The organization of the 30 nm fibre has been studied by X-ray diffraction[10], electron microscopy[7,11,12], neutron scattering[13,14] and electric and flow dichroism[8,15–18], with general agreement on a somewhat irregular solenoidal helix of pitch about 11 nm, diameter 30 nm, and approximately six nucleosomes per helical turn. Further structural details have, however, been elusive, and divergent models incorporating a 'superbead' rather than a solenoidal structure have even been proposed[19].

The linear dichroism of the 30-nm fibre is sensitive to the orientation of linker DNA and nucleosomal disks relative to the fibre axis[8,15,18]. However, since the optical dichroism is an average of linker and disk contributions, it has not previously been possible to deduce unambiguously the orientation of either of these structural elements. In this article we describe measurement of the photochemical dichroism for psoralen–DNA adduct formation in naked DNA and chromatin fibres. Since psoralen can be shown to react primarily with the linker DNA, we are able to calculate the separate dichroism contributions of linker and disk, and hence to estimate their orientations relative to the fibre axis.

Photochemical dichroism

Since a photochemical reaction requires an initial electronic excitation event, the angular orientation of the corresponding transition moment can be determined by measuring the photochemical yield as a function of polarization direction relative to an external orientation axis. Our experiment utilizes the photochemical reaction of aminomethyltrioxsalen (AMT)[20] with DNA or chromatin. After dark equilibration of the solution with ^{3}H-labelled AMT, a brief electric field pulse (0–40 kV cm^{-1}) is applied across the sample, and, with allowance for a delay time to permit orientation, the solution is flashed with a polarized 308-nm pulse from an excimer laser (Fig. 1). The extent of covalent attachment of psoralen to adjacent pyrimidines is assayed by scintillation counting of the purified DNA sample.

The optical arrangement with a rotatable polarizer allows us to use light polarized either parallel or perpendicular to the electric field direction. The photochemical yields, normalized for variations in the laser pulse, are $Y_{\parallel}$ and $Y_{\perp}$ respectively. The photochemical dichroism ρ_{pc} at field E is computed from the ratios

$$R_{\parallel} = Y_{\parallel}(E)/Y_{\parallel}(E=0)$$

$$R_{\perp} = Y_{\perp}(E)/Y_{\perp}(E=0)$$

with

$$\rho_{pc}(E) = \frac{3(R_{\parallel} - R_{\perp})}{(2R_{\perp} + R_{\parallel})}$$

Measuring both parallel and perpendicular polarizations allows one to correct for isotropic nondichroism effects such as dissociation of a bound chromophore in the electric field.

Orientation of psoralen in naked DNA

To determine the orientation of linker DNA segments in chromatin, it is essential that we know the orientation of the psoralen excitation moment relative to the local double helix axis. For this purpose we incubated ^{3}H-AMT with high-molecular weight herring sperm DNA at a ratio of 1 AMT per 200 bp; a single 308-nm laser pulse was sufficient to yield several hundred counts per minute above background from these solutions. We also measured the 266-nm optical DNA dichroism of the samples, and, with an increase of the AMT to DNA ratio to 1 per 10 bp, we were able to determine the absorbance dichroism for psoralen intercalated into DNA. The results summarized in Fig. 2 show that AMT photochemical dichroism matches closely to both the UV dichroism of the bases, and to the optical dichroism of intercalated AMT.

We conclude that the psoralen transition moment responsible for photochemical reaction, and the DNA base transition moments, lie at nearly the same angle to the helix axis. This is the result expected for a π-π^* excitation transition lying in the plane of an intercalated psoralen molecule. The photochemical data are not, however, sufficiently precise to distinguish the ideal angle of nearly 90°, anticipated for intercalation in a classical B-form helix, from the perceptibly tilted (~75°) angle found for the base transition moments in short DNA fragments[21].

AMT binding in 30-nm chromatin fibre

Even though earlier reports suggested a strong preference of psoralen for the chromatin linker region in nuclei[22,23], quantitative interpretation of the photodichroism experiment obliges us to measure directly the relative binding to linker and nucleosomal DNA in the isolated, Mg^{2+}-compacted chromatin fibre. Chromatin samples irradiated in the presence of AMT were

© 1984Nature Publishing Group

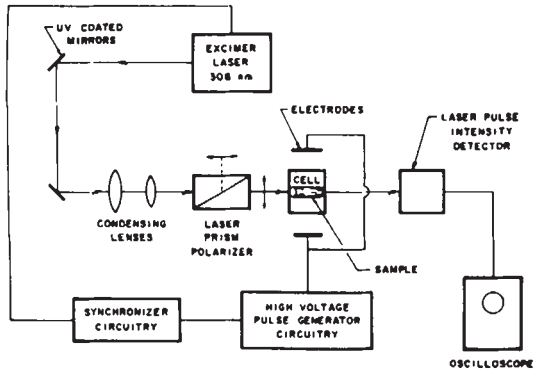


Fig. 1 Schematic representation of the photodichroism apparatus. A pulsed excimer laser (Lambda Physik 201) operating with xenon chloride gas mixtures was used to irradiate the samples at 308 nm. The high voltage electric field is generated via a capacitative discharge across two platinum electrodes in a temperature jump cell modified by filling the cell with an agarose gel, except for the cylindrical optical path between the quartz windows[28]. In this way the entire sample can be irradiated. The high-voltage pulse was generated by capacitor discharge, triggering at the same time a pulse generator which, after a typical preset delay time of 10–100 μs (depending on the orientation time of the sample), triggers the laser. The intensity of each laser pulse is monitored using a laser intensity detector and a Tektronix storage oscilloscope. This allowed us to correct for any variation in laser intensity in our final calculation of photochemical yields. The laser beam is polarized either parallel or perpendicular to the applied electric field by use of a rotatable UV-quality Glan Thomson prism polarizer used for high-power laser operations. The preset delay times between the electric field and the laser pulse were checked using a fast response photodiode and a storage oscilloscope.

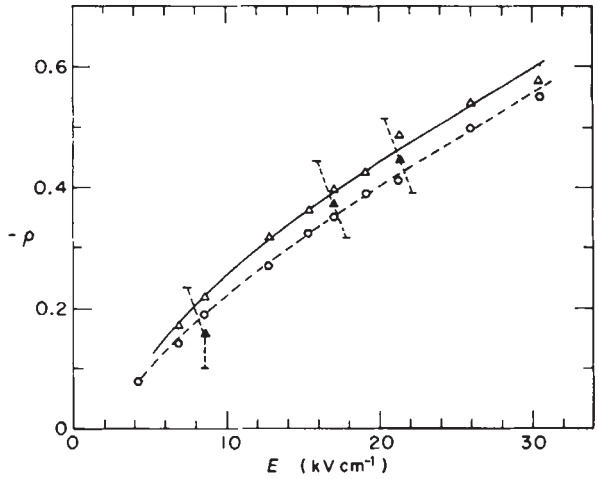


Fig. 2 Summary of the optical dichroism and the photochemical dichroism experiments on high molecular weight herring sperm DNA. ○, Reduced dichroism values for herring sperm DNA observed at the wavelength of base pair absorption, 266 nm. △, reduced dichroism values for the dark complex of AMT intercalated in DNA at a ratio of AMT to base pairs of 1:10. The absorption changes were monitored at 311 nm where AMT has considerable absorption but the base pairs do not absorb. ▲, Photodichroism values obtained for the covalent attachment of AMT to herring sperm DNA.

Methods: Herring sperm DNA (Sigma) was phenol extracted, ethanol precipitated, dissolved in 0.5 mM Tris *p*H 8.0, 0.05 mM EDTA and dialysed extensively against the same buffer. For the photodichroism experiment, DNA at $A_{260}=0.25$ was equilibrated with tritiated AMT in the dark in the ratio of 1 AMT molecule to 200 bp. A 320 μl sample was irradiated in the dichroism cell, and after thorough mixing in the cell, a 200 μl aliquot was removed, and the DNA was ethanol precipitated in the presence of yeast RNA three times to eliminate any non covalently bound AMT. The final precipitate was dissolved in 150 μl 10 mM Tris, 1 mM EDTA, *p*H 8 buffer and counted in a Packard 3320 Tri-Carb scintillation counter.

digested with micrococcal nuclease, and the resulting DNA was subjected to agarose gel electrophoresis (Fig. 3). Gel photographic negatives showing ethidium fluorescence were scanned to quantitate the amount of DNA in each band; careful control of exposure conditions is required for linear film response, which was monitored with control samples containing known amounts of DNA. Bands were excised from the gel and counted for ^{3}H-AMT content. Nucleosome dimers (360 ± 30 bp) contained an average of 2.2 times as much ^{3}H-AMT per base pair as found for nucleosomes trimmed to remove the linker. We calculate from this result that the 36 bp linker DNA accounts for 60 ± 10% of the AMT binding under the conditions of our experiment, and the 166-bp nucleosome accounts for the remaining 40 ± 10%.

Dose/response in AMT–chromatin interaction

Analysis of the photodichroism experiment requires that the extent of photochemical attachment of AMT to DNA be a linear function of the amount of radiation absorbed. Figure 4*a* shows a verification of this dependence at a ratio of 1 AMT added per 200 bp. We also were concerned that added AMT might perturb chromatin structure. Figure 4*b* shows that the amount of covalent attachment is linear in amount of AMT added, as long as the levels of added AMT are kept below 1/25 bp. Hence there is no evidence for a cooperatively induced chromatin structural change upon AMT binding. In addition, we found (data not shown) that addition of AMT to 1/200 bp does not affect the measured optical dichroism of the 30-nm fibre, either before or after photoreaction.

Photochemical dichroism of chromatin fibres

We measured the photochemical dichroism for the AMT-chromatin reaction in calf thymus fibres stabilized by addition of 0.5 mM Mg^{2+}, in 0.1 mM Tris, 0.01 mM EDTA, *p*H 7.5. These conditions were chosen because earlier experiments showed that the chromatin fibre electrical orientation function behaved in the same way as observed for fibres highly cross-linked to prevent electric field-induced distortion dichroism[15,16,18]. When nondichroism effects (absorbance changes) are removed, it is found that field-induced orientation follows a classical induced dipole mechanism, with saturation of orientation at fields above about 10 kV cm^{-1} (see Fig. 6). The increasingly negative dichroism reported by McGhee *et al.*[8] at high fields may be due to field-induced distortion dichroism of the fibre, possibly due to use of lower Mg^{2+} (0.1 mM) concentrations. Field-induced distortion is an especially difficult problem for chromatin fibres because of the large permanent dipole moment (1,200 D) of the nucleosome[24]; the experiments of McGhee *et al.*[8,25] do not persuasively eliminate complications due to fibre distortion.

Figure 5 shows the dichroism results. The photochemical dichroism is substantially more negative than the optical dichroism, and is, within experimental error, constant over the range 10 kV cm^{-1} to 17 kV cm^{-1}. The mean of these values is $\rho_{pc}=-0.23\pm0.04$. The nondichroism correction, measured as the extent of departure from unity of the quantity $(2R_\perp+R_\parallel)/3$ is within 3% for measurements up to 15 kV cm^{-1} field strength. Above that field, it is of the order of −10%. Figure 5 also shows results for the optical dichroism, along with comparative values from our earlier studies. We confirm the value of −0.09 for the optical dichroism of Mg^{2+}-stabilized fibres; fully crosslinked fibres have positive optical dichroism ($\rho=0.05$).

Dichroism from nucleosome and linker

We are now in a position to separate the dichroism contributions of linker and nucleosome, subject to the underlying assumption that photochemical reaction of AMT can occur anywhere along the linker length with equal probability, and that the AMT

© 1984 Nature Publishing Group

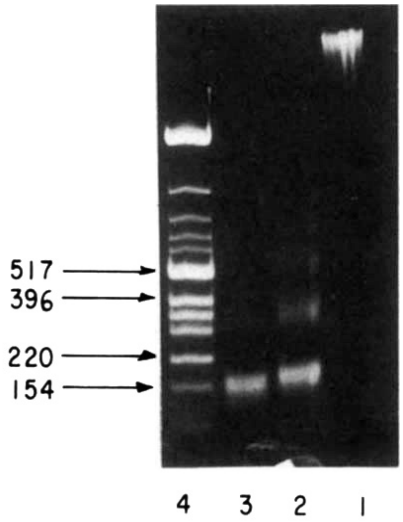


Fig. 3 Micrococcal nuclease digestion of chromatin fibres irradiated in the presence of AMT. The chromatin fibres were incubated with AMT at a ratio of AMT to base pairs of 1:200. The irradiation buffer was 0.1 mM Tris, *p*H 7.5, 0.01 mM EDTA and 0.5 mM Mg^{2+} and the irradiation source for this experiment was a broad spectrum UV source generating 300–400 mm. Following irradiation the sample was made up to 15 mM Tris *p*H 7.5, 1.2 mM $CaCl_2$ and digested micrococcal nuclease for various times at 37 °C. The digestion products were proteased, phenol extracted and the resulting DNA alcohol precipitated several times before running on a 2.5% agarose gel. The ladder pattern in lane 2 is after 30 s and lane 3 after 2 min of nuclease digestion. Lane 4 contains the *Hin*fI digest of pBR322. Lane 1 shows whole chromatin DNA before nuclease digestion. The amounts of AMT covalently bound in each band are shown in Table 1.

Methods: Gel bands were cut out and the radioactivity in each band measured. A negative of the ethidium fluorescence from the gel was scanned using a Zenith soft laser scanning densitometer (Biomed.) and the area under each peak was determined as a proportional measure of DNA in each band. Control samples of known relative DNA content were used to ensure linearity of film response (data not shown). Quantitative comparison of these results leads us to the conclusion that 60% ± 10% of AMT binding is to the linker DNA. Chromatin fibres for these and for all our photodichroism experiments were prepared by gentle staphylococcal nuclease digestion of calf thymus nuclei[15]. The fibres were fractionated by size in a 10–30% sucrose gradient in 10 mM Tris, 1 mM EDTA, 60 mM NaCl. We present data on fractions 40–80 nucleosomes in length, or 8–16 kb DNA. We have also determined the photodichroism of chromatin fibres 20–60 nucleosomes in length, but the results remain the same as those presented here and in Table 1 for fibres 40–80 nucleosomes in length. The DNA sizes were checked on 1% agarose gels with a *Hae*III digest of λDNA as size markers (not shown). The presence of H1 and the core histones were routinely checked by 15% acrylamide gels containing 0.1% SDS. The fibres were judged free of RNA contamination as no absorbance changes were observed after RNase treatment.

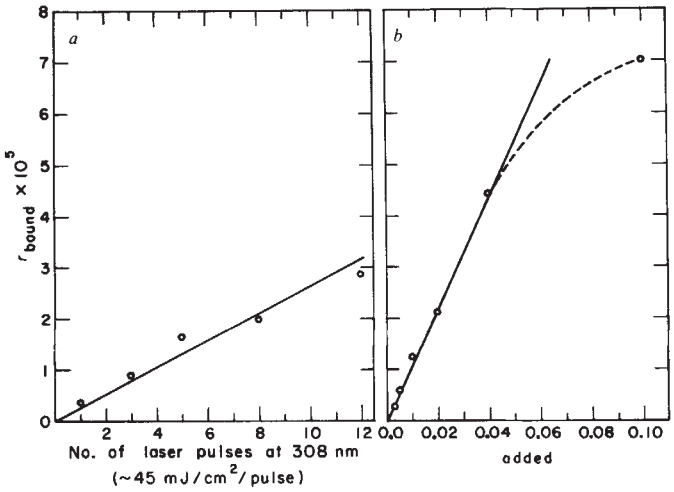


Fig. 4 *a*, Dose dependence of the covalent binding of AMT to the 30 nm chromatin fibre. Sample fibres 40–80 nucleosomes long (8–16 kb DNA) prepared as described in Fig. 3 were incubated at an AMT to base pair ratio of 1:200. Samples were irradiated with laser pulses at 308 nm. Following irradiation each sample was deproteinized using proteinase K and the DNA was alcohol precipitated several times to remove noncovalently bound psoralen. The amount of AMT attached covalently is shown as the ratio bound per base pair. *b*, Covalent binding of AMT to the magnesium stabilized chromatin fibre in 0.1 mM Tris, 0.01 mM EDTA, 0.5 mM Mg^{2+}, *p*H 7.5, $T = 20$ °C. The ratio of AMT to base pairs was varied, but the total amount of laser irradiation at 308 nm was left constant for each sample. The amount of AMT covalently attached was determined as described earlier and is expressed as the ratio bound per base pair of DNA in chromatin.

Table 1 The amount of AMT covalently bound in each band on the gel shown in Fig. 3

Particle	^3H counts (c.p.m.)	Amount of DNA (arb. units)	Counts/ area	Particle size (bp)
Monomer	1145	13.2	87	180±15
Dimer	590	4.5	131	360±30
Partially trimmed monomer	565	9.3	61	150−165
Linkerless dimer	64	1.1	58	310±5

transition moment is therefore cylindrically averaged about the linker DNA helix axis. In addition, we use the results of Fig. 2 to justify interchanging photochemical and UV optical dichroism values.

Linker DNA constitutes 0.17 (34/200) of the total absorbance, and accounts for 0.6 of the AMT binding. Hence

$$0.17\rho_l + 0.83\rho_n = -0.09 \quad (1)$$

$$0.6\rho_l + 0.4\rho_n = -0.23 \quad (2)$$

where ρ_l and ρ_n are respectively the dichroism of linker and nucleosomal DNA; the first equation is for the optical dichroism, and the second for the photochemical experiment. The solution of these equations is $\rho_l = -0.36 \pm 0.07$; $\rho_n = -0.04 \pm 0.03$. The uncertainty estimates include the probable error in the fraction of AMT binding to the nucleosome.

Orientation of nucleosomes and linker DNA

We assume that the 166 bp nucleosome contains an integral number of half superhelical turns of DNA. The dichroism is then closely approximated by[24]

$$\rho_n = 3/8(3\cos^2\gamma - 1) \quad (3)$$

where γ is the angle formed between the fibre axis and the normal to the plane of the nucleosome disk (equivalent to the uperhelix axis of the nucleosomal DNA). We find $\gamma = 57° \pm 3°$, where the uncertainty limits include allowance for possible tilting of the AMT excitation moment by the same extent as found for the DNA base transition moments in rod-like DNA[24]. Our earlier paper details the modifications of equation (3) required to account for such effects[24]. Stated in an alternative way, the results show that the nucleosomal disk is tilted on average by 33 ± 3° away from the solenoid axis (Fig. 6).

The path of linker DNA is less certain than is the case for the nucleosome. If we assume a semicircular arc for the linker, the angle describing its orientation can be found from equation (3). The result is $\gamma = 83° \pm 7°$, equivalent to a tilt angle of 7° ± 7°. In such a model, the linker DNA local superhelix axis is thus found to be nearly perpendicular to the solenoid axis; alternatively, the plane containing the semicircular arc is tilted by only 7° away from the solenoid axis. Another model for the linker

© 1984 Nature Publishing Group

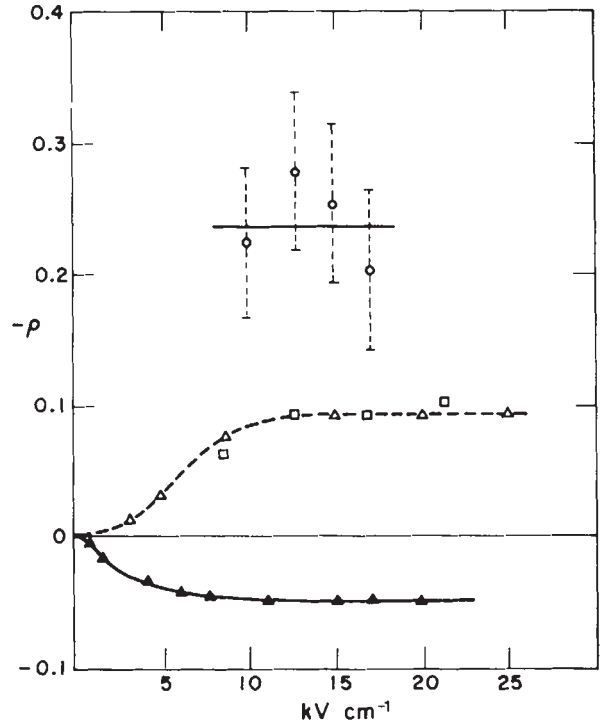


Fig. 5 Summary of the results of the photochemical and optical dichroism experiments on Mg^{2+} stabilized chromatin fibres and dimethyl-suberimidate crosslinked chromatin fibres. ○, Values for the photochemical dichroism obtained for the AMT chromatin complex at different applied electric fields. □, △ Optical dichroism of Mg^{2+} stabilized calf thymus chromatin at different electric fields as determined by two sets of independent experiments. ▲, Optical dichroism values of dimethylsuberimidate crosslinked fibres obtained earlier by Yabuki *et al.*[18].
Methods: The error bars in photochemical dichroism data represent 67% confidence limit on the numbers. The chromatin fibres were obtained from calf thymus glands as described earlier and were stabilized in 0.5 mM Mg^{2+}, 0.1 mM Tris, 0.01 mM EDTA, *p*H 7.5. The ratio of AMT to base pairs was 1:200. Five laser pulses at 308 nm produced sufficient covalent attachment for counting each sample at a particular polarization with or without the electric field. The chromatin was treated as described in Fig. 3 to measure the amount of psoralen bound, $T = 20\,°C$.

is a straight rod. The dichroism is then closely approximated by[24]

$$\rho = -3/4(3\cos^2\beta - 1)$$

where β is the angle between the rod and the solenoid axis. Using $\rho = -0.36$ yields $\beta = 45° \pm 3°$.

Implications for chromatin structure

Uncertainty about the topology of DNA in chromatin prevents formulation of a definitive model at this time. However, our results do provide concrete values for the orientation of nucleosomes and linker DNA in the fibre, allowing us to evaluate the compatibility of general classes of models with our observations. The 33° angle found for the tilt of the nucleosomal disk-faces relative to the solenoidal axis eliminates models in which the disks are either parallel[8] or perpendicular[14] to the solenoid axis.

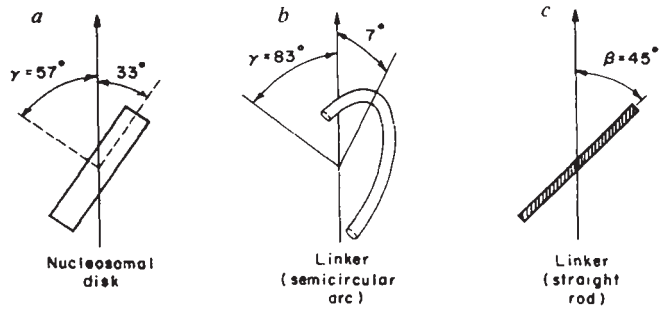


Fig. 6 Schematic representation of the orientation of the nucleosome disk and the linker DNA in calf thymus chromatin as deduced from our photodichroism work. In *a*, γ represents the angle formed between the solenoid axis and the normal to the plane of the nucleosome disk and in *b*, γ represents the angle between the linker DNA superhelix axis and the solenoid axis where the assumption is that the linker is a semicircular arc. In *c*, β is the angle between the linker DNA rod and the solenoid axis where the linker DNA is assumed to be a rod. The solenoid axis is represented by the vertical line with an arrow.

Also of general significance is the substantial difference between linker and nucleosome dichroism. This rules out models in which DNA in chromatin is assumed to follow the same superhelical path in nucleosome and linker, with a constant number of base pairs per superhelical turn. Because of variations in linker length, the linker in such models must lose its phasing relative to the DNA superhelix, and must therefore occupy a variable position relative to the solenoid axis[25]. Averaged over many turns, the dichroism of linker and nucleosome will be the same for such models. Our results clearly require that the linker have a special structure, or a regular position in the fibre, or both. Other specific models, such as the non-solenoidal twisted ribbon of Worcel *et al.*[26], can also be ruled out because their predicted linker dichroism is more positive than we measured. Possible models are discussed systematically elsewhere[27].

Work is currently in progress in our laboratory on the photochemical dichroism of chromatin of different linker lengths. In addition, we are attempting to determine the photochemical dichroism for DNA chain cleavage reactions in order to elucidate the intrinsic DNA periodicities in chromatin fibres. We expect the photochemical dichroism method to be applicable to many questions relating to the organization of nucleic acids in cellular subunits and viruses.

This work was supported by NIH grant GM 21966 and grant 61-582 from the Jane Coffin Childs Memorial Fund for Medical Research. S.M. is a fellow of the Jane Coffin Childs Memorial Fund for Medical Research.

Received 1 July; accepted 22 December 1983.

1. Simpson, R. T. *Biochemistry* **17**, 5524–5531 (1978).
2. Finch, J. T. *et al. Nature* **269**, 29–36 (1977).
3. Finch, J. T., Levit-Bentley, A., Bentley, G. A., Roth, M. & Timmins, P. A. *Phil. Trans. R. Soc.* **290**, 635–638 (1980).
4. Bentley, G. A., Finch, J. T. & Levit-Bentley, A. *J. molec. Biol.* **145**, 771–784 (1981).
5. Olins, A. L. & Olins, D. E. *Science* **183**, 330–332 (1977).
6. Kornberg, R. D. *A. Rev. Biochem.* **46**, 931–954 (1977).
7. Thoma, F., Koller, T. & Klug, A. *J. Cell. Biol.* **83**, 403–427 (1979).
8. McGhee, J. D., Rau, D. C., Charney, E. & Felsenfeld, G. *Cell* **22**, 87–96 (1980).
9. Finch, J. T. & Klug, A. *Proc. natn. Acad. Sci. U.S.A.* **73**, 1897–1901 (1976).
10. Sperling, L. & Klug, A. *J. molec. Biol.* **112**, 253–263 (1977).
11. Olins, A. L. & Oline, D. E. *J. Cell. Biol.* **81**, 260–265 (1979).
12. Rattner, J. B. & Hamkalo, B. A. *Chromosoma* **69**, 373–379 (1978).
13. Hollandt, H., Notbohm, H., Riedel, F. & Harbers, E. *Nucleic Acids Res.* **6**, 2017–2027 (1979).
14. Suau, P., Bradbury, E. M. & Bladwin, J. P. *Eur. J. Biochem.* **97**, 593–602 (1978).
15. Lee, K. S., Mandelkern, M. & Crothers, D. M. *Biochemistry* **20**, 1438–1445 (1981).
16. Lee, K. S. & Crothers, D. M. *Biopolymers* **21**, 101–116 (1982).
17. Tjernfeld, F., Norden, B. & Wallin, H. *Biopolymers* **21**, 343–358 (1982).
18. Yabuki, H., Dattagupta, N. & Crothers, D. M. *Biochemistry* **21**, 5015–5020 (1982).
19. Renz, M., Nehls, P. & Hozier, J. *Proc. natn. Acad. Sci. U.S.A.* **74**, 1879–1883 (1977).
20. Isaacs, S. T., Shen, C. J., Hearst, J. E. & Rapoport, H. *Biochemistry* **16**, 1058–1064 (1977).
21. Hogan, M., Dattagupta, N. & Crothers, D. M. *Proc. natn. Acad. Sci. U.S.A.* **75**, 195–199 (1978).
22. Hanson, C. V., Shen, C.-k. J. & Hearst, J. E. *Science* **193**, 62–64 (1976).
23. Wieschahn, G. P., Hyde, J. E. & Hearst, J. E. *Biochemistry* **16**, 925–932 (1977).
24. Crothers, D. M., Dattagupta, N., Hogan, M., Klevan, L. & Lee, K. S. *Biochemistry* **17**, 4525–4533 (1978).
25. McGhee, J. D., Nickol, J. M., Felsenfeld, G. & Rau, D. C. *Cell* **33**, 831 (1983).
26. Worcel, A., Strogatz, S. & Riley, D., *Proc. Nat. Acad. Sci.* **78**, 1461 (1981).
27. Mitra, S., Sen, D. & Crothers, D. M. *J. Biomolec. Struct. Dynam.* (submitted).
28. Riesner, D., Coplan, M. & Randles, J. W. *Analyt. Biochem.* **121**, 186–189 (1982).

© 1984Nature Publishing Group

Radio detection of a jet in the Crab Nebula

T. Velusamy

Radio Astronomy Centre, Tata Institute of Fundamental Research, PO Box 8, Ootacamund 643 001, India

The jet-like feature in the Crab Nebula discovered by van den Bergh[1] shows a tubular structure of 75×45 arc s with strikingly well-collimated sharp parallel boundaries in the recent deep O III image obtained by Gull and Fesen[2]. I now report detection of radio emission at 20 cm from this faint optical jet, in the Very Large Array (VLA) map obtained with a resolution of 15×15 arc s. The radio emission is extremely faint at less than 1% of the peak brightness within the nebula, and its detection was possible because of the high dynamic range in this VLA map. Its positional coincidence and structural similarity to the optical jet are remarkable. The emission at 20 cm appears to be nonthermal, with strong linear polarization of ~30%.

The optical jet in the Crab Nebula was first discovered in a deep exposure Schmidt plate as a peculiar jet-like feature extending to about 100 arc s beyond the northern boundary of the Crab Nebula[1] and is seen best in Fig. 1, which is a high contrast display of the image, reproduced from Gull and Fesen[1]. This jet roughly coincides with a spur in the outermost contour of Woltjer's[3] optical continuum map of the Crab. However, this feature is most visible in [O III] λ 5,007 emission[4,5]. Its emission is sharply defined over a rectangular area of size 75×45 arc s. The jet appears to be limb brightened with astonishingly straight edges. It is located about 3 arc min north of the pulsar, but is not radially aligned with the pulsar. It has low radial velocity 150 km s^{-1}, suggesting that it is oriented roughly in the plane of the sky. No radio emission has been reported from the jet previously because of lower dynamic range in earlier maps[6,7], although there is some indication for it in the Westerbork Synthesis Radio Telescope (WSRT) maps at 21 and 49 cm (refs 8, 9).

The 20 cm VLA maps of the Crab Nebula presented here were made from a total of 6 hours of observations in the C-configuration, during April 1983. A bandwidth of 12.5 MHz was used at 1.4 GHz. The phase centre used for the observations was at the position of the Crab pulsar. 3C138 and 3C147 were used for all calibrations including linear polarization. Data were analysed following usual VLA reduction procedure. All the maps were cleaned and restored with a beam of 15×15 arc s. A dynamic range close to about 1,000 was achieved in the final self-calibration map.

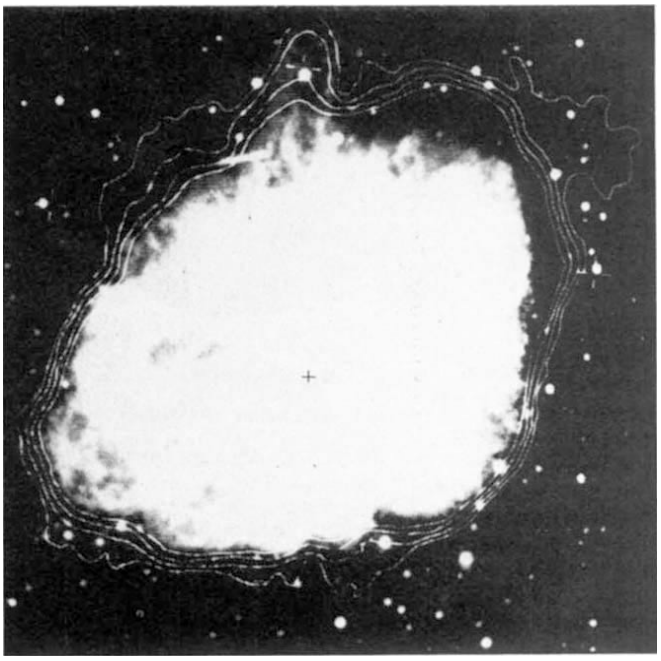


Fig. 1 Selected low level radio contours of the 20-cm VLA map superimposed on the very high contrast O III image obtained by Gull and Fesen[1]. The contours are 12, 24, 60 and 120 mJy per beam. The radio peak is 6,336 mJy per beam. The cross marks the pulsar.

Figure 1 shows an overlay of a few low-level radio contours of the Crab Nebula superimposed on the optical O III image and Fig. 2 shows detailed maps of the total and polarized intensities at 20 cm. The detailed structures seen in these maps at very low brightness levels (<50 mJy per beam) particularly over the optical jet are very remarkable, considering the presence of very strong emission of 1,000 to 6,000 mJy per beam within the nebula. Due to small distortions in the images in Fig. 1, it was not possible to align the optical and radio to an accuracy better than 5 arc s. In general, there is very good agreement between the optical and radio boundaries except in the north-west where the radio emission extends well beyond the nebular boundary in O III. It may be noted that the O III boundary is very sharp over this part of the nebula. Also, in the 20-cm VLA map there is no evidence for any circumnebular radio shell around the Crab, as has been suggested by the optical and X-ray data[10,11]. An upper limit to the surface brightness of such a

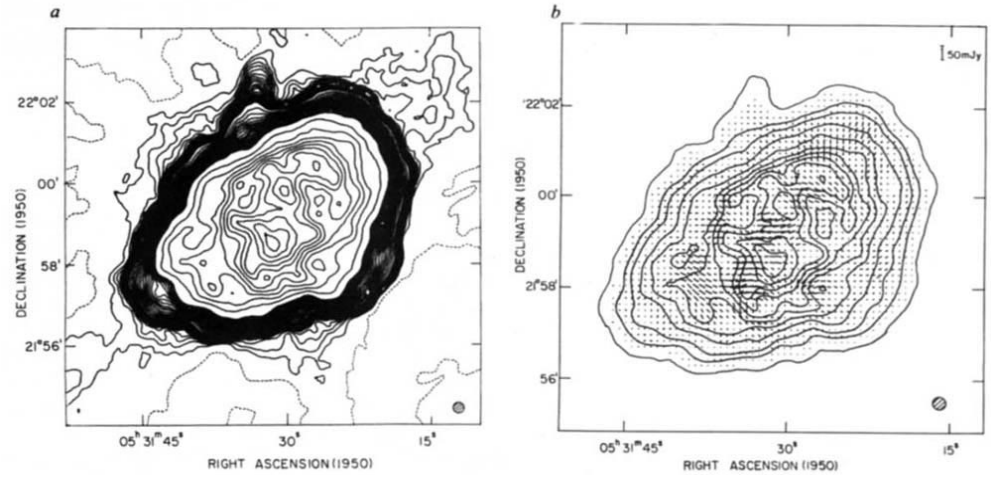


Fig. 2 *a*, VLA map of the Crab Nebula at 20 cm with a resolution of 15×15 arc s. The hatched circle represents the size of the synthesized beam. The lowest and highest contours are 5 and 6,000 mJy per beam. Negative values are indicated by dashed contours. The contours shown are at three intensity ranges at intervals given within parentheses, as follows: 10 to 100 (5); 100 to 1,000 (50) and 1,000 to 6,000 (500) mJy per beam. *b*, The distribution of linearly polarized intensity. The intensity and position angle of the E-vector of the polarized emission are indicated by the length and orientation of the vectors.

© 1984 Nature Publishing Group

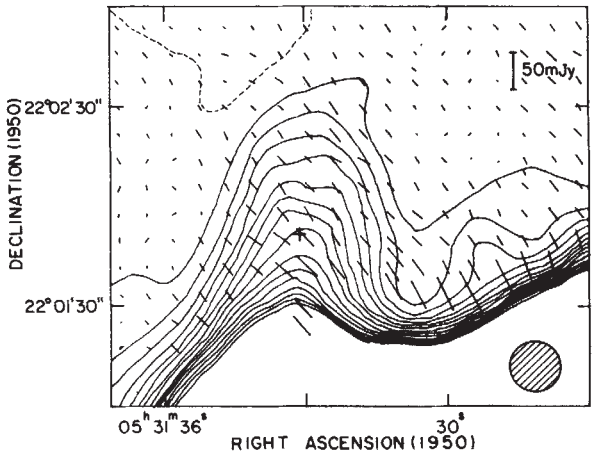


Fig. 3 An enlargement of the jet in the VLA map at 20 cm. The distribution of intensity, and position angle of the linearly polarized emission are represented by the vectors. The contours represent the distribution of the total intensity. The lowest contours is 6 mJy per beam and the contour interval is 6 mJy per beam.

shell in the VLA map is $\sum_{1.4\,\mathrm{GHz}} < 0.6\times10^{-20}\,\mathrm{W\,m^{-2}\,Hz^{-1}\,sr^{-1}}$; the earlier estimate is $\sum_{0.6\,\mathrm{GHz}} < 2\times10^{-20}\,\mathrm{W\,m^{-2}\,Hz^{-1}\,sr^{-1}}$ by Wilson and Weiler[8]. A detailed discussion of the intensity, and polarization structure within the nebula, as well as the circumnebular emission are presented elsewhere.

In Fig. 2, the jet is clearly seen at brightness less than 50 mJy per beam, which is less than 1% of the peak emission in the nebula. As can be seen from Fig. 2*b*, the jet is also visible in the polarized intensity map. Figure 3 shows an enlargement giving structural details of the jet in the total and polarized intensities. The polarization data is noisier than the total intensity, possibly because of the inaccuracies in the instrumental polarization. The radio structure is consistent with a sharply bounded rectangular area of size about 75×45 arc s, which is the same as the optical extent. There is good agreement between the overall structure of the jet in the optical and radio regions (Figs 1, 2). The limb brightening along its walls seen in the optical jet is not obvious in Fig. 3, possibly because of the lower resolution. The strong polarization of 30–50% in Fig. 3 suggests that the 20 cm emission from the jet is nonthermal. Also, there is some indication for the presence of the jet in the WSRT 49-cm map[8] at a level of 850 to 1,300 mJy arc min^{-2}, while the mean brightness in the VLA 20-cm map is ~480 mJy (arc $\mathrm{min})^{-2}$. Thus the radio spectrum appears to be nonthermal with a rough estimate for the spectral index between 49 and 20 cm of the order of −0.8 or less.

The strong polarization as well as the distribution of the orientation of E-vector indicates a highly ordered magnetic field along the nebular boundary and in the jet. Assuming a value of $-25\,\mathrm{rad\,m^{-2}}$ for the Faraday rotation in the nebula close to the jet[6] we can get some idea of the projected magnetic field pattern over the jet and in the nebula. The orientation of the projected magnetic field appears to be at position angles between 0° and 20°. That is, the magnetic field is approximately along the direction of the jet and there is no appreciable change in its orientation between the nebula and the jet. It has been suggested that the jet consists of some unusually high velocity ejecta from SN 1054 or similar, but separate explosion[2]. Some radio emission may be expected from the shock front associated with the ejecta. However, the observed radio structure (Fig. 3) shows no evidence for any such shock front near the northern end of the jet. On the other hand, its radio surface brightness, $\sum_{1.4\,\mathrm{GHz}} \sim 7\times10^{-20}\,\mathrm{W\,m^{-2}\,Hz^{-1}\,sr^{-1}}$, limb brightening and magnetic field structure, seem to indicate that the radio emission originates in the interstellar matter and magnetic field swept up along the walls of the jet as in the case of Kundt's model[13]. It is therefore likely that the jet may show some radial expansion.

Is there any other radio spur or jet present in the Crab nebula? An examination of the low brightness (<20 mJy per beam) features in Fig. 2*a*, suggests at least two possible cases. The extended features north-west and south-east along the major axis may not be real since the strongest sidelobes were in this direction. However, the broad feature (~2 arc min) along the northeastern boundary seems to be real; it is also noticeable in polarized intensity map. The presence of the feature along the southern boundary approximately diametrically opposite to the prominent northern jet is very interesting. If real, it would indicate that the jets are the result of symmetric injection of particles from the pulsar.

It has been suggested that the jet is a tubular trail of material lost by the progenitor of the pulsar[12] or a result of plasma instabilities near the outskirts of the nebula[4,13,14]. Chevalier and Gull[4] have suggested that the jet is indeed a thin gaseous sheet fanning radially outward from dense interior filaments. The sheets may be created by differential acceleration or by Rayleigh–Taylor instability. In the models of Kundt[13] and Bychkov[14], the high velocity (~0.1*c*) streams of relativistic particles from the pulsar lead to instabilities producing (1) 'corridors' at the filament-plasma boundary, passing through the filament and out into the interstellar medium, or (2) a 'chimney' caused by particles pushing against the ambient heavier interstellar medium. The jet in the Crab Nebula may be such a corridor or chimney or a tubular cavity left behind by the progenitor through which some plasma mixed with relativistic particles is flowing outward. The radio structure, spectrum and magnetic field pattern indicated by the polarization, seem consistent with the above interpretations. In that case, a variation in the radio spectrum along the jet may be observable. Clearly, multifrequency observations with higher resolution will be required to understand fully the origin of such jets and their radio emission.

I thank Drs Mohan Joshi and Pramesh Rao for useful comments and R. A. Fesen for the O III photographs. I am grateful to Professor M. R. Kundu for hospitality offered during the observations. The VLA of National Radio Astronomy Observatory is operated by Associated Universities Inc. under contract with NSF.

Received 1 November; accepted 8 December 1983.

1. van den Bergh, S. *Astrophys. J. Lett.* **160**, L27 (1970).
2. Gull, T. R. & Fesen, R. A. *Astrophys. J. Lett.* **260**, L75–L78 (1982).
3. Woltjer, L. *Bull. astr. Inst. Neth.* **13**, 301 (1957).
4. Chevalier, R. A. & Gull, T. R. *Astrophys. J.* **200**, 399–401 (1975).
5. Davidson, K. *Astrophys. J.* **228**, 179–190 (1979).
6. Wilson, A. S. *Mon. Not. R. astr. Soc.* **157**, 229–253 (1972).
7. Swinbank, E. & Pooley, G. *Mon. Not. R. astr. Soc.* **186**, 775–778 (1979).
8. Wilson, A. S. & Weiler, K. W. *Nature* **300**, 155–157 (1982).
9. Duin, R. M. & van der Laan, H. *Astrophys. Lett.* **12**, 177–180 (1972).
10. Murdin, D. & Clark, D. H. *Nature* **294**, 543–544 (1981).
11. Toor, A., Palmeri, T. M. & Seward, F. D. *Astrophys. J.* **207**, 96–101 (1976).
12. Blandford, R. D., Kennel, C. F., Mckee, C. F. & Ostriker, J. P. *Nature* **301**, 586–587 (1983).
13. Kundt, W. *Astr. Astrophys.* **121**, L15–L18 (1983).
14. Bychkov, K. V. *Sov. Astr.* **18**, 420–425 (1975).

On the energy balance of solar active regions

Gary A. Chapman

San Fernando Observatory, California State University, Northridge, California 91330, USA

The cause of sunspots has long been an important, unsettled problem in solar physics. Biermann[1] suggested that the strong magnetic field of a sunspot inhibited convection, allowing the sunspot to cool. Parker[2], on the other hand, proposed that a sunspot was cooled by the generation of waves that carried away the thermal energy. The solar 'constant' has been measured with a daily uncertainty of about 10 parts per million (p.p.m.) by the Active Cavity Radiometer Irradiance Monitor (ACRIM) on board the Solar Maximum Mission (SMM) satellite launched in February 1980. The ACRIM has shown that

© 1984 Nature Publishing Group

the solar constant changes with solar activity, showing dips of the order of 0.1–0.3% associated with sunspots. This discovery has provoked renewed attempts to understand the sunspot phenomenon. Two divergent views are emerging: that the missing energy is stored in the convection zone for long periods of time[3] or that the missing energy is re-radiated fairly quickly at different angles by faculae[4], non-spot magnetic fields that often accompany sunspots. The consequences are that in the first case an 11-yr modulation is expected in the solar output whereas in the second case it is not. I point out here that facular emission may equal the missing energy from sunspots, over a period of some weeks or months, suggesting that the missing energy is stored by being converted from kinetic into magnetic energy after which it decays back into thermal energy in the faculae.

The problem with the Oster model of facular re-radiation is that for some sunspots, most notably those seen in April 1980, no substantial facular area was associated with these regions. To see the magnitude of the problem we need only look at the irradiance deficit associated with the April 1980 disk passage of active region, Boulder number 2370/2372, which was responsible for an irradiance drop of about 0.16% at maximum. Assuming a triangular form for the time-dependent irradiance deficit, we have an energy deficit of $(\sim 1.2\times10^{30}\ \mathrm{erg\ s^{-1}}\times 12\ \mathrm{days}) = 1.2\times10^{36}$ erg. Because there was no significant facular presence associated with this active region, we are forced to conclude that much of this energy was stored in the Sun.

We know that large sunspots decay over a period of days, weeks or even months, producing large facular regions (showing as bright plage in the chromosphere). Willson *et al.*[5] have already suggested that faculae may later re-radiate the energy that was blocked by sunspots. If this is true it implies that the missing energy is stored in some fashion in the magnetic fields of the active region. I now present evidence that this may indeed be the case and briefly discuss the implication for the nature of sunspots and magnetic fields in solar active regions.

In addition to the arguments presented by Oster, Schatten and Sofia[4], based primarily on statistical models, the following approximate calculation is offered. The fractional flux change due to sunspots or faculae can be written as $\Delta F/F_0 = \int_0^1 (\Delta I/I_0)\cos\theta\, d\omega$, where θ and ω are the heliocentric and the azimuthal angles as seen from the Earth.

Let us assume that

$$\frac{\Delta I}{I} \simeq \text{const.} = -0.5 \qquad (1)$$

and

$$\frac{\Delta I}{I} = b(\mu^{-1} - a) \qquad (2)$$

for sunspots and faculae, respectively[6], where a and b are empirically determined constants and $\mu = \cos\theta$. Then for sunspots we have that $\Delta F/F = \pi(\Delta I/I) = -0.5\ \pi$, and for faculae we have $\Delta F/F \simeq \pi b$ $(a = 1)$. If we let $b = 0.03$, then $\Delta F/F = 0.03\ \pi$. Thus, we see that, per unit area, faculae could re-radiate 6% of the energy blocked by sunspots.

Independent evidence has shown that the facular magnetic flux density is 1,500–1,800 G or about half of that in a typical, large sunspot. If we assume that the sunspot and facular region form a bipole, then the magnetic flux in the sunspot must be equal in magnitude to that in the facular region, suggesting that faculae have about twice the area of their associated sunspots. Thus, faculae, per unit time, can radiate as much as $2\times6\% = 12\%$ of the missing flux of a sunspot. Therefore, if faculae can last as much as $(0.12)^{-1}$ times longer than the sunspot we might have a balance in the total energy. According to Allen[7] the typical lifetime of sunspots is 6–11 days, whereas for plage (the chromospheric manifestation of magnetic fields) the lifetime is about 43 days. Thus, the ratio of lifetimes is 4–8. Foukal[8] found a value of 83 days for the lifetime of an active region. Harvey[9] determined 90 days for the average lifetime of active regions from He I 10830 spectroheliograms. The point is that, based on simple models, faculae can radiate from half to all of the energy blocked by sunspots. It should be emphasized that this re-radiation is not immediate but has a delay of some days between the maximum area of the sunspots and the maximum area of the faculae.

The fluctuations in the ACRIM data have been compared with simple models for facula and sunspot irradiance variation. The models are

$$\mathrm{PSI} = -C_s A_s\ \mu(3\mu + 2) \qquad (3)$$

and

$$\mathrm{PFI} = C_p A_p\ (\mu - 3\mu^2 + 2) \qquad (4)$$

where PSI is a photometric sunspot index, PFI is a photometric facular index and C_s and C_p are empirically determined constants. These indices represent changes in solar irradiance in p.p.m. if the areas of spots and plage (A_s and A_p) are expressed in millionths of a hemisphere. In this form, Hudson's PSI would require a value of 0.164 for C_s (ref. 5). The form of equation (3) is like that of Hudson's PSI. The value of PFI goes to zero at disk centre ($\mu = \cos\theta = 1$).

We have selectively examined several periods in 1980 when the ACRIM showed very low and very high levels of solar irradiance. The time periods were day of year 98–111, 130–142, 147 and 199–214. For these 35 days of data the extreme ACRIM data values were 1,369.34 to 1,366.28 W m^{-2}.

A multiple regression was carried out with $C_s = 0.109$ and $C_p = 0.0388$ in equations (3) and (4) of the form:

$$\Delta\mathrm{ACRIM} = A + B\ \mathrm{PSI} + C\ \mathrm{PFI} \qquad (5)$$

where $\Delta\mathrm{ACRIM} = (\mathrm{ACRIM}/1{,}368.24 - 1)\times10^6$. The quantity ΔACRIM expressed the fluctuation in solar irradiance in p.p.m. from the 1980 average value.

The values obtained for the multiple regression coefficients are $A = 261$, $B = 1.3075$ and $C = 0.6725$. The multiple correlation coefficient R was 0.937 and the standard error of estimate was 220 p.p.m. Linear regressions between any two of the three data sets gave lower correlation coefficients. More importantly, the values of the two multiple regression coefficients A and B do not differ from unity by as much as a factor of two, and are not more than about a factor of two from each other, implying that facular emission cannot be ignored.

If we use the results of the above regression to redetermine the coefficients in equations (3) and (4) we find that $C_s = 0.143$ and $C_p = 0.0261$.

The areas and positions for the spots and faculae used in equation (5) were obtained from the Solar Geophysical Data Bulletin. The results of equation (5) suggest that both faculae and sunspots are needed for a good fit. Circular shape of facular (or plage) regions leads to a value for C_p of $(4/\pi)\times0.03 \simeq 0.04$. If we assume a filling factor for faculae, in a plage, of 0.10 (ref. 10), this value of C_p is consistent with the value assumed in equation (10) by Chapman[6], which referred to an individual facula.

A more recent analysis (G.A.C. and A. D. Meyer, in preparation), referring to 45 active regions, observed by the Extreme Limb Photometer (ELP)[6], gives a value for C_p in equation (4) of 0.0092 ± 0.0014 $(t = 6.5)$. (It must be emphasized that this value corresponds to the area of the calcium plage in millionths of a hemisphere.) The value of $C_s = 0.164 \pm 0.0083$ $(t = 12.8)$. The data have not been modified by bolometric corrections[6], because they are approximately equal and opposite to corrections for stray light.

Calculating the energy balance of faculae and sunspots, we have the time-averaged flux deficit of sunspots,

$$\left(\frac{\Delta F}{F}\right)_s = 2\pi \int_0^1 (\mathrm{PSI})\, d\mu = -4\pi C_s A_s \qquad (6)$$

and the time-averaged excess of faculae,

$$\left(\frac{\Delta F}{F}\right)_f = 2\pi \int_0^1 (\mathrm{PFI})\, d\mu = 3\pi C_p A_p \qquad (7)$$

© 1984 Nature Publishing Group

The ratio of these two quantities is $(\Delta F/F)_f/(\Delta F/F)_s = 1.4$, where we have taken $A_p/A_s \sim 25$ (ref. 11). This result even more strongly suggests that there is an energy balance between sunspots and faculae.

Thus, photometric observations, obtained with the ELP, support the view that faculae can, on the average, radiate away the deficit flux associated with sunspots. This conclusion must be checked by further observations as well as by further comparisons with satellite irradiance monitors.

If energy balance exists between sunspots and faculae on a time scale of months, one would not expect to see an 11-yr modulation in the solar output due to active regions. If an 11-yr variation is observed, it will imply that there are global causes of solar variance, perhaps changes in convective efficiency[12].

If energy balance exists between sunspots and faculae as is argued here, where is the energy stored, since facular regions outlast sunspots by a large factor of 8–14 (ref. 7)? It is difficult to avoid the conclusion that the energy must be stored in the magnetic fields of the active region itself. Since it is hard to imagine how the magnetic fields could store thermal energy, we suggest that the energy is stored as magnetic energy. This idea has already been proposed by P. R. Wilson (personal communication and refs 13, 14) and, perhaps, others.

A sunspot is the result of conversion of kinetic into magnetic energy and the faculae are the liberation of this energy back into thermal energy. Stenflo has suggested that faculae may be heated by resistive dissipation[15]. For the scale of faculae or filigree, classical resistivity is too low and we must invoke some form of anomalous resistivity. Work on solar flares gives examples of these processes (see ref. 16 for a recent review).

For the phase of the radiative disturbance to agree with the generation of magnetic fields, the diameter and flux of a sunspot flux tube must reach a maximum within a few days and subsequent conversion of kinetic to magnetic energy must result in growth of the length of the flux tube.

The depth of the flux tube of an active region can be estimated from the separation rate of the leader and follower poles. The separation rate of the leader and follower poles[17] is 0.2 km s^{-1}. Assuming a growth time of 45 days, this separation rate results in a maximum depth of $\simeq 0.5\ R_0$. This large depth suggests that sunspots are very deep phenomena transversing the entire depth of the convection zone. The energy to drive this 'engine' may come from organized motions such as the torsional oscillator[18]. A more realistic scenario might be the concurrent generation and destruction of magnetic fields, as facular regions are often observed to develop around large sunspots shortly after their appearance.

I acknowledge useful conversations with Drs J. Lawrence, H. Hudson and R. Willson and thank Mr A. D. Meyer for help in obtaining preliminary results from ELP data. This research was supported in part by NSF grant AST-8121863.

Received 22 August; accepted 1 December 1983.

1. Biermann, L. *Vierteljahrsschr. astr. Ges.* **76**, 194 (1941).
2. Parker, E. N. *Sol. Phys.* **36**, 249–274 (1974).
3. Eddy, J. A., Gilliland, R. L. & Hoyt, D. V. *Nature* **300**, 689–693 (1982).
4. Oster, L., Schatten, K. H. & Sofia, S. *Astrophys. J.* **256**, 768–773 (1982).
5. Willson, R. C. *et al. Science* **211**, 700–702 (1981).
6. Chapman, G. A. *Astrophys. J. Lett.* **242**, L45–L48 (1980).
7. Allen, C. W. *Astrophysical Quantities* 3rd edn (Athlone, London, 1973).
8. Foukal, P. V. in *The Physics of Sunspots* (eds Cram, L. E. & Thomas, J. H.) 391–423 (Sacramento Peak Observatory, Sunspot, 1981).
9. Harvey, J. in *Workshop on Solar Irradiance Variations on Active Region Time Scales* Pasadena (eds La Bonte, B., Chapman, G. A., Hudson, H. S. & Willson, R. C. (NASA Conference Proceedings, Washington DC, in the press).
10. Chapman, G. A. in *Solar Active Regions* (ed. Orrall, F. Q.) 43–82 (Colorado Associated University Press, Boulder, 1981).
11. Foukal, P. & Vernazza, J. *Astrophys. J.* **234**, 707–715 (1979).
12. Dearborn, D. S. P. & Blake, J. B. *Astrophys. J.* **237**, 616–619 (1980).
13. Wilson, P. R. in *The Physics of Sunspots* (eds Cram, L. E. & Thomas, J. H.) 83–97 (Sacramento Peak Observatory, Sunspot, 1981).
14. Wilson, P. R. preprint, Univ. Sydney (1983).
15. Stenflo, J. O. *IAU Colloq.* **36**, 143 (1977).
16. Priest, E. in *Solar Active Regions* (ed. Orral, F. Q.) 213–275 (Colorado Associated University Press, Boulder, 1981).
17. Sheeley, N. R. Jr in *Solar Active Regions* (ed. Orral, F. Q.) 17–42 (Colorado Associated University Press, Boulder, 1981).
18. Howard, R. & La Bonte, B. J. *Astrophys. J. Lett.* **239**, L33–L36 (1980).

Solar gravity modes as a test of turbulent diffusion mixing

G. Berthomieu, J. Provost & E. Schatzman

Observatoire de Nice, B.P. 139, 06003 Nice Cedex, France

The Stanford group[1] has recently detected solar global oscillations in the range 160–370 min and interpreted them as internal gravity modes of degree $l=1$ and $l=2$. Other observations[2,3] also indicate the existence of low-degree solar gravity modes. Whereas the high-degree 5-min oscillations are sensitive to the properties of the convective zone essentially, the low-frequency modes (gravity modes) are sensitive to the physical conditions in the solar core. Other constraints on the solar core are given by the observed value of the neutrino flux, which is lower than the flux predicted by the standard models, and by the splitting of the low-degree 5-min oscillations. As far as the neutrino flux is concerned, a possibility to reduce it has been explored by Schatzman and Maeder[4] by introducing a turbulent diffusion mixing. In this paper we study the influence of such mixing on the periods of low-degree l gravity modes, which do not present the difficulties of the high-degree gravity modes[5]. We find that these periods depend strongly on the pseudo-Reynolds number[4], and we derive the range of this parameter compatible with the recent observational results.

The basic assumption of Schatzman and Maeder is that some sort of marginal instability generates turbulence. The turbulent transport flattens the distribution of the elements in the solar core, modifies the usual balance between the various nuclear processes and leads to a reduction of the central temperature. The turbulent diffusion coefficient is assumed to be given by the law $D_T = R_e^* \nu$, where ν is the molecular viscosity and R_e^* the pseudo-Reynolds number. Models with different values of R_e^* have been computed by Schatzman and Maeder and their properties relative to the neutrino flux and solar surface abundances of different chemical elements are discussed in ref. 4.

The low-frequency gravity modes can be studied with an asymptotic method[6]. For a solar-type model with an inner radiative zone and a convective envelope the periods are given at the lowest order by:

$$P = \frac{P_0}{\sqrt{l(l+1)}}\left(k + \frac{l}{2} - \frac{1}{4}\right) \tag{1}$$

with

$$P_0 = \frac{2\pi^2}{\int_0^{x_c} \frac{N}{c}\,dx} \tag{2}$$

where N is the Brunt–Väissälä frequency, x_c the limit between the radiative interior and the convective envelope, k is the number of radial nodes. This asymptotic formula is derived for the low degree and high radial order modes, and neglecting the perturbation of the gravitational potential (Cowling approximation). Equation (1) implies that low frequency gravity modes with a given l have equidistant periods. The spacing period P_0 depends only on the Brunt–Väissälä frequency N in the inner radiative zone. This expresses the fact that g-modes are driven by the buoyancy force which is locally measured by N. In fact, a more realistic solar model has an external radiative zone, and the phase factor $(l/2-1/4)$ in equation (1) should be modified by a small amount which can be estimated of the order of 10^{-2}–5×10^{-2} for small l, $k<50$ and assuming an effective surface polytropic index of three. This correction remains small except when there is a resonance between g-modes associated with the two radiative zones[7]. Such a resonance does not occur for the range of periods we have considered here, since the periods of g-modes associated to the external radiative zone are much larger. P_0 has been calculated for the three models of

© 1984 Nature Publishing Group

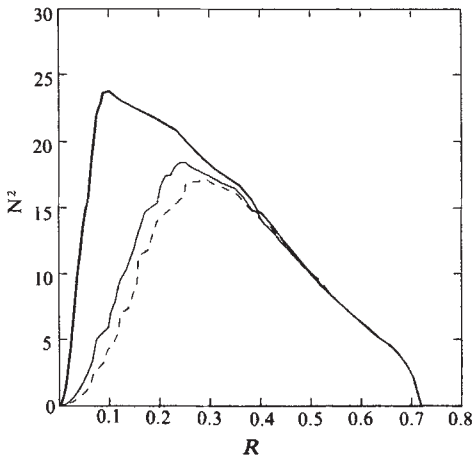


Fig. 1 Square of Brunt–Väissälä frequency plotted against radius for the three models. ——$R_e^*=0$; ——$R_e^*=100$; - - - - $R_e^*=200$.

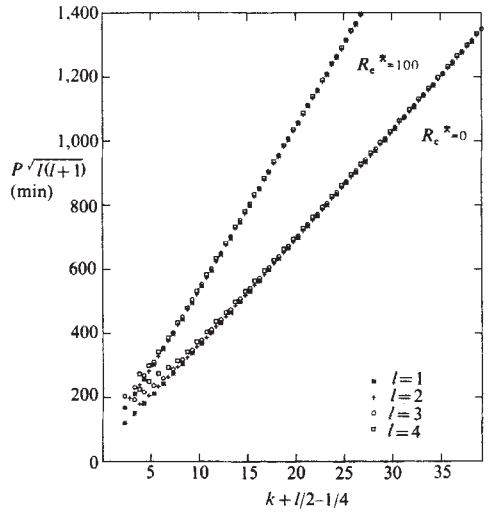


Fig. 2 Periods of gravity modes multiplied by $\sqrt{l(l+1)}$ plotted against $k+l/2-1/4$ (k is the radial order and l the degree) for the two models $R_e^*=0$ and $R_e^*=100$. Note that for sufficiently high radial order, the asymptotic relation is a good approximation.

Schatzman and Maeder with a chemical composition hydrogen, $X=0.73$; heavy elements, $Z=0.02$; their main characteristics and the periods P_0 are given in Table 1. From Table 1 it appears that increasing values of R_e^* correspond both to a decrease of the neutrino flux and to an increase of the spacing period P_0. As a consequence of the diffusion, the temperature gradient increases, the gradient of the mean molecular weight decreases due to the flattening of the hydrogen distribution and to the enlargement of the ^{3}He peak. Thus the Brunt–Väissälä frequency decreases as can be seen in Fig. 1 where its values have been plotted versus the radius for the three models $R_e^*=0$, 100, 200.

As the asymptotic formula has been derived at the limit of very low frequencies and with some assumptions concerning the model and the oscillations, we have tested its validity by calculating numerically the periods of the linear adiabatic gravity modes with $l=1$, 2, 3, 4 and radial order $k=2,\ldots,40$ for the two models $R_e^*=0$ and $R_e^*=100$. The perturbation of gravitational potential is taken into account. The results are plotted in Fig. 2. For each model and for sufficiently high values of k ($k>10$ or 15) the points are distributed along a straight line. This means that the behaviour of the oscillations of a realistic model is in good agreement with the asymptotic formula. The slope of the line gives another estimation of P_0: P_0^*. Restricting to the modes

Table 1 Characteristic properties of the three models of Schatzman and Maeder

R_e^*	$\log L/L_\odot$	$\log R/R_\odot$	N_ν (SNU)	P_0	$\bar{P}_0$
0	0.045	0.0028	11,96	34,3	33,9
100	0.007	−0.0136	2,33	52	54,6
200	−0.006	−0.0202	1,43	58,7	62,9

Data shown are luminosity, radius, neutrino flux and values (in min) of the spacing period P_0, given by equation (2) and the renormalized corresponding values $\bar{P}_0$.

Table 2 Estimated spacing periods P_0 for models with different initial solar abundances

Model	X	Z	N_ν (SNU)	P_0 (min)
Iben *et al.*, 1976				
Model I	0.817	0.01	4.4	38
Model II	0.744	0.02	10.3	35
Model III	0.6745	0.03	24.7	32
Christensen-Dalsgaard *et al.*, 1979				
Model A	0.73	0.02	5.2	36
Model B	0.81	0.004	2.3	38
Model C	0.84	0.001	1.7	37.7

with $k>20$ we obtain by a least-square analysis:

$$P_0^*=34.4 \text{ min with } \sigma=0.2 \text{ min for model } R_0^*=0$$

$$P_0^*=52.5 \text{ min with } \sigma=0.15 \text{ min for model } R_e^*=100$$

σ being the root-mean square difference between computed periods and their asymptotic approximations. These values of P_0^* are in approximate agreement with the values of P_0 derived from the equation (2) which are given in Table 1. This agreement shows that neglecting the perturbation of the gravitational potential has little influence on the value of P_0. The three solar models do not have exactly the standard solar radius. This has been taken into account by multiplying the values obtained for P_0 by a factor $(R/R_\odot)^{-3/2}$. The renormalized values $\bar{P}_0$ of P_0 are given in the last column of Table 1.

A relation between the values of the spacing period $\bar{P}_0$ and the value of the corresponding pseudo-Reynolds number of the turbulent diffusion can be derived by an interpolation formula. An analytical study of the diffusion equation leads to an eigenvalue problem. It turns out that the departure from the zero-diffusion model varies by $(1-\exp(R_e^*\langle\nu\rangle t^*/L^2)$, where t^* is the age of the star, $\langle\nu\rangle$ an average of the kinematic viscosity and L some characteristic length which includes the solution of the eigenvalue problem. As this latter quantity varies with the gradient of the density, the departure from the zero-diffusion model will vary with R_e^* according to the same law. We then write,

$$N^2_{R_e^*}=N^2_{R_e^*=0}+a(1-e^{-bR_e^*})$$

where a and b are r-dependent coefficients which will be derived from the models. We shall assume by a simple dimensional analysis that it is legitimate to write for the spacing period $\bar{P}_0$ an interpolation formula of the same kind,

$$(\bar{P}_{0,R_e^*})^{-2}-(\bar{P}_{0,R_e^*=0})^{-2}=-A(1-e^{-BR_e^*})$$

With the values calculated for $\bar{P}_{0,0}$, $\bar{P}_{0,100}$ and $\bar{P}_{0,200}$ given in Table 1 we obtain: $A=6.34\times10^{-4}\,\text{min}^{-2}$, $B=1.87\times10^{-2}$. The variation of $\bar{P}_0$ according to this law is given in Fig. 3. We note that for $R_e^*<100$ the variation of $\bar{P}_0$ is almost linear with a slope of the order of $\partial\bar{P}_0/\partial R_e^*\sim0.2$ min.

It is interesting to compare the effects of diffusive mixing with the effects of chemical composition changes. A variation of the spacing period relative to the initial abundances of hydrogen X and heavy elements Z can be roughly derived from published models of Iben *et al.*[8] and Christensen-Dalsgaard *et al.*[9]. For

© 1984 Nature Publishing Group

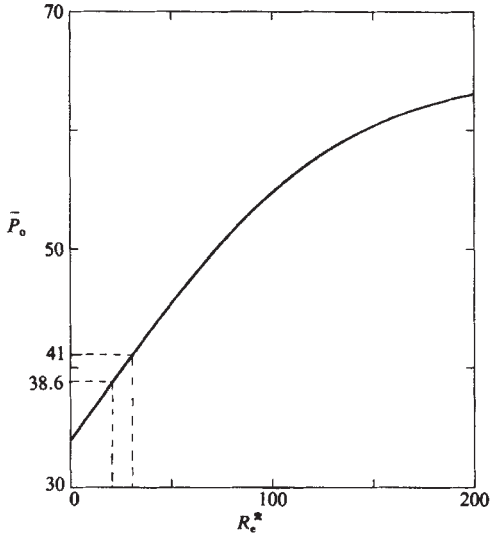


Fig. 3 Variation of the spacing period $\bar{P}_0$ relatively to the pseudo-Reynolds number R_e^*. Dashed lines indicate observed values.

these models, the neutrino flux N_ν and the period P_0 estimated from the published values of the periods of the g-modes are reported in Table 2. These results show that a variation of the initial abundances which lowers the neutrino flux increases the value of the spacing period, but in a weaker way than a variation of R_e^*, since a chemical variation which reduces the neutrino flux by a factor 2.5 increases the period of about 2 or 3 mn while a mixing which reduces the neutrino flux by 2.6 ($R_e^* = 50$) increases P_0 of about 10 mn.

The interpretation of the observations by the Stanford group gives a period spacing P_0 of 38.6 ± 0.5 min. From the Solar Maximum Mission/ACRIM (Active Cavity Radiometer Irradiance Monitor) data with frequencies greater than 40 μHz, Froehlich *et al.* derived a period spacing of 38.9 min. However, applying a statistical method to a larger range of data (down to 10 μHz), they do not obtain a single period spacing but indications of a rather high P_0 (ref. 22). With periods P_0 of the order of 34–36 min, standard models do not fit these results. The period P_0 of some low Z and low helium models are closer to the observed values but these models have a convective zone too shallow to be compatible with the p-modes eigenfrequencies of high and low degree[10–12]. Our results show that turbulent diffusion-mixed models with quite moderate values of the pseudo-Reynolds number ($R_e^* \sim 25$) are able to give the observed periods. Let us see now how such models are compatible with the other constraints.

A solar model must nowadays satisfy an increasing number of constraints. The present investigation has been carried under the assumption that a consistent solar model can be built with just one more assumption: the presence of turbulent diffusion. A recent discussion of the generation of turbulence by shear flow in rotating stars[13] suggests a time-dependent turbulent diffusion, and it is quite clear that complete time dependent solar models should be studied. As a first approximation we have assumed that the extra assumption mentioned above is a stationary turbulent diffusion coefficient given by the relation: $D_T = R_e^* \nu$, where ν is the kinematic viscosity (in the case of the Sun, it is 90% molecular viscosity). The choice of R_e^* is supposed to be possible in such a way as to fulfil the constraints concerning the splitting of low-degree pressure modes, the surface abundance of lithium, the surface abundance of ^{3}He, the solar neutrino flux and the period spacing of the g-modes.

The observations of high-degree 5-min oscillations favour models with a large convective zone[14–16]. This constraint is not incompatible with turbulent diffusion mixing since the depth of the convective zone is very little dependent on the mixing. Concerning the splitting between the frequencies of low-degree five-minute oscillations with radial order differing by one and degree differing by two, that is $\nu_{k,l} - \nu_{k-1,l+2}$, we obtained by asymptotic analysis that it is multiplied by about a factor 5/3 from $R_e^* = 0$ to $R_e^* = 100$. This is in agreement with the results of Ulrich *et al.*[11] which used a mixed model built with the hydrogen distribution given by Schatzman and Maeder for $R_e^* = 100$, and they found a splitting too large to account for the observations. However, for $R_e^* = 25$, the factor is reduced to 1.17 which means that the splitting is not much modified by such moderate diffusion.

The surface abundance of ^{3}He, ^{3}He/^{4}He $\simeq 5 \times 10^{-4}$ according to Geiss and Reeves[17] has to be split into two parts, one coming from the formation of ^{3}He from initial ^{2}D, the other one carried by turbulent diffusion from the ^{3}He peak in the solar interior. With the present estimates of ^{2}D (ref. 18) the ^{3}He abundance provided by turbulent diffusion mixing is of the order of (^{3}He/^{4}He) $\simeq 3 \times 10^{-4}$, which implies a maximum value of R_e^* of the order of 40. Then comes the question of the lithium depletion. The convective zone needed to account for the properties of the high-frequency oscillations has a depth of about 200,000 km, corresponding to a temperature of the order of 2.15×10^6 K. The lithium burning region at $T \sim 2.5 \times 10^6$ K sits in a depth of about 230,000 km. A simple model of turbulent mixing from $T = 2.15 \times 10^6$ K to $T = 2.5 \times 10^6$ K leads to a value of R_e^* of about 40 (with $\nu_{mol} \sim 15$). A slight penetration of the convection[13] of about one tenth of the scale height leads to a value of R_e^* of the order of 25.

The neutrino flux produced by a model with R_e^* of the order of 25 is about 0.6 times the flux produced by the standard model. This is still too large even if the comparison with other standard models[19,20] with a lower N_ν suggests that it is possible to reduce the flux significantly. Thus it seems difficult to bring this flux to the level of the last value of Davis[21], $N_\nu = 2.2 \pm 0.4$ SNU. It is quite clear from these numbers that the simple model with R_e^* constant throughout the whole Sun seems to present a contradiction between values of R_e^* derived from the neutrino flux and from the oscillations.

The value of R_e^* derived from the period spacing of the g-modes, $R_e^* \simeq 25$, seems to be compatible with the other constraints, except the constraint on the neutrino flux. It should be borne in mind however that an actual model is not an exact model, that all its characteristics should be adjusted to fit the different constraints and that a more sophisticated treatment of turbulent diffusion should be used. The main conclusion is that the model with zero diffusion mixing does not seem to fit the present interpretations of the observations of low-frequency oscillations (g-modes) whereas a model with moderate diffusion mixing and the proper chemical composition seems to fulfil the g-modes, the ^{3}He surface abundance and the ^{7}Li depletion and is compatible with the high-frequency oscillations constraints.

We thank P. Delache for stimulating discussion and communication of results prior to publication, A. Maeder for providing the solar models and D. Gough for critical comments on the manuscript.

Received 10 October; accepted 28 November 1983.

1. Delache, P. & Scherrer, P. *Nature* **306**, 651–654 (1983).
2. Van der Raay, H. B. *Oscillations as a Probe of the Sun's Interior* (European Study Conference, Catania, 1983).
3. Froelich, C. & Delache, P. *Oscillations as a Probe of the Sun's Interior* (European Study Conference, Catania, 1983).
4. Schatzman, E. & Maeder, A. *Astr. Astrophys.* **96**, 1–16 (1981); *Nature* **290**, 683–686 (1981).
5. Gough, D. O. *Pleins Feux sur la Physique Solaire* (ed. Rösch, J.) (CNRS, 1978).
6. Tassoul, M. & Tassoul, J. L. *Ann. Astrophys.* **31**, 251–256 (1968).
7. Tassoul, M., *Astrophys. J. Suppl. Ser.* **43**, 469–490 (1980).
8. Iben, I. & Mahaffy, J. *Astrophys. J.* **209**, L39–L43 (1976).
9. Christensen-Dalsgaard, J., Gough, D. O. & Morgan, J. G. *Astr. Astrophys.* **73**, 121–128 (1979).
10. Christensen-Dalsgaard, J. & Gough, D. O. *Nature* **288**, 544–547 (1980).
11. Ulrich, R. K. & Rhodes, E. J. *Astrophys. J.* **265**, 551–563 (1983).
12. Gough, D. O. *Nature* **302**, 18 (1983).
13. Zahn, J. P. *Astrophysical Processes in Upper Main Sequence Stars* (eds Cox, A. N., Vanclair, S. & Zahn, J. P.) 253–329 (13th Advanced Course Swiss Society of Astronomy and Astrophysics, SAAS-FEE, 1983).
14. Ulrich, R. K. & Rhodes, E. J. *Astrophys. J.* **218**, 521–529 (1977).
15. Berthomieu, G. *et al. Lect. Not. Phys.* **125**, 307–312 (1980).
16. Lubow, S. H., Rhodes, E. J. & Ulrich, R. K. *Lect. Not. Phys.* **125**, 300–306 (1980).
17. Geiss, J. & Reeves, H. *Astr. Astrophys.* **18**, 126–132 (1972).
18. Peimbert, M. *Proc. Workshop on Primordial Helium and Related Light Elements*, Garching (European Space Observatory, 1983).
19. Bahcall, J. N., Huebner, W. F., Lubow, S. H., Parker, P. D. & Ulrich, R. K. *Rev. mod. Phys.* **54**, 767–799 (1982).

© 1984 Nature Publishing Group

20. Christensen-Dalsgaard, J. *Mon. Not. R. astr. Soc.* **199**, 735–761 (1982).
21. Davis, R. Jr *Proc. Brookhaven Solar Neutrino Conf.* Vol. 1 (BNL 50879, 1978).
22. Froehlich, C. & Delache Ph. *Proc. Solar Seismology from Space, Snowmass 1983* (Jet Propulsion Laboratory, Pasadena, 1984).

First known EL5 chondrite—evidence for dual genetic sequence for enstatite chondrites

Derek W. G. Sears & Karen S. Weeks

Department of Chemistry, University of Arkansas, Fayetteville, Arkansas 72701, USA

Alan E. Rubin*

Department of Mineral Sciences, National Museum of Natural History, Smithsonian Institution, Washington, DC 20560, USA

The enstatite chondrites form two compositionally distinct groups—the EH and EL chondrites—and four petrologic types (types 3–6), where each type represents increasing degrees of metamorphic alteration[1,2]. Prior to the recovery of the Antarctic meteorite Reckling Peak A80259, the subject of the present report, all known EL chondrites were petrologic type 6, whereas all known EH chondrites were types 3–5. It has long been appreciated that the variations in bulk composition preclude a simple conversion of EH4 material into EL6 material[3–7]. However, complex models involving simultaneous variations in bulk composition and petrologic type have been discussed[1,5,8,9] and may be implied by other classification schemes in common use; that of Anders-Keil (type I–intermediate–type II) and that of Van Schmus and Wood (E3-E4-E5-E6). We report here the discovery of the first EL5 chondrite. An EL5 breaks the EH3,4-EH5-EL6 sequence, and indicates that the enstatite chondrites constitute two discrete, isochemical metamorphic sequences, EH3–5 and EL5–6.

The enstatite chondrites are a small (25-member) class of meteorites which formed in a uniquely reducing environment. As a consequence, they contain many terrestrially unknown minerals, such as niningerite ((Mg,Fe)S), and Si-bearing metal. They are extremely heterogeneous and show large variations in their bulk composition; sulphur and other chalcophile elements vary by a factor of three and iron and the other siderophiles vary by a factor of approximately two. They also differ from each other in the extent to which they have suffered metamorphism; some are only slightly recrystallized while others are highly recrystallized, possess coarse textures and have experienced complete obliteration of primary structures such as chondrules[3–6]. When these characteristics were first described, it seemed as though there was a single sequence of enstatite chondrites. At one end of the sequence were meteorites which Anders[4] and Keil[6] termed type I, with high siderophile element abundances and with textures displaying small amounts of recrystallization. These graded through two intermediate meteorites (St Sauveur and St Marks) to type II, with low siderophile element abundances and highly recrystallized textures. The situation was apparently very different from the ordinary chondrite case, where textural variations occur in each of the three coherent chemical groups.

In 1967 Van Schmus and Wood[1] introduced a highly successful two-dimensional grid scheme for classifying meteorites in which bulk composition constitutes one dimension and extent of metamorphic alteration (petrologic type) constitutes the other. Keil's type I, intermediate type and type II are equivalent to E4, E5 and E6 in this scheme, even though the E class thus defined is not isochemical. Sears *et al.* suggested that the E class actually comprises two isochemical classes, EH and EL, and emphasized that the EH and EL chondrites were not simply related by the sequence EH4-EH5-EH6. The discovery of an EL5 chondrite now provides strong evidence that the enstatite chondrites are not a single genetic sequence.

The Reckling Peak A80259 meteorite is a highly weathered, 20.2 g stone which was found in the Antarctic during the 1980–81 summer season. B. Mason originally classified it as an E5 chondrite[10]. We have carried out a detailed petrologic examination of the meteorite using the library section at the Smithsonian Institution, and we have performed a bulk analysis using neutron activation analysis. Modal analyses (Table 1) were made microscopically, using an automated point counter; mineral vol. % was converted into wt % using estimated mineral densities. Mineral compositions (Table 2) were determined with an electron microprobe using crystal spectrometers and following standard Bence–Albee and ZAF correction procedures. Cobalt was determined in Fe,Ni metal after subtracting the contribution of the Fe K_β peak from Co K_α. Two 150 mg fragments of Reckling Peak A80259 were analysed by instrumental neutron activation analysis, along with two other enstatite chondrites and a variety of primary synthetic and secondary standards. The data reported here were gathered in four irradiations with duplicate chips always being run in

Table 1 Modal abundance (wt %) in Reckling Peak A80259

Enstatite	47
Plagioclase	4
Kamacite	10
Limonite	28
Troilite	10
Niningerite	0.5
Schreibersite	<0.1
Total	99.5
Points counted	1,555
Area analysed (mm^2)	143

Table 2 Mineral compositions (wt %) in Reckling Peak A80259

	Enstatite		Kamacite	Troilite	Niningerite
No. of grains	16		25	26	12
SiO_2	59.4	Si	2.1		
Al_2O_3	0.19	Ni	6.4		
FeO	0.37	Co	0.43		
MgO	39.9	P	0.64		
CaO	0.69	Fe	89.7	58.1	32.1
MnO	<0.08	Mg		0.07	11.2
Total	100.55	Ca		<0.04	3.0
End member	$Fs_{0.5}Wo_{1.2}$	Mn		0.48	12.8
		Cr		3.1	1.3
		Ti		0.47	<0.08
		Zn		0.10	<0.09
		S		37.8	41.5
		Total	99.27	100.12	101.9

* Present address: Institute of Geophysics and Planetary Physics, University of California at Los Angeles, Los Angeles, California 90024, USA.

© 1984 Nature Publishing Group

separate irradiations. Irradiations were performed at the University of Missouri research reactor and samples were counted for various times over a period of 6 weeks.

Reckling Peak A80259 consists predominantly of enstatite, kamacite and troilite, with minor plagioclase and accessory niningerite and schreibersite (Table 1). It contains approximately 12 vol.% radial, barred and porphyritic pyroxene chondrules, 980–1,600 μm in apparent diameter. Chondrule textures are recrystallized but readily delineated. Euhedral enstatite grains exhibit highly undulose extinction and range in size from 15 to 375 μm. They occur in kamacite- and/or troilite-rich areas and in apparently recrystallized areas surrounded by anhedral enstatite and plagioclase. Approximately 90% of the enstatite exhibits parallel extinction under crossed polarizers and, hence, is probably orthorhombic. Because clinoenstatite may also exhibit parallel extinction under certain crystal orientations, the actual abundance of clinoenstatite is somewhat greater than 10%. Plagioclase occurs as very small (≤2 μm) grains, interstitial to euhedral enstatites. Plagioclase volatilized very rapidly under electron bombardment and was impossible to analyse with the microprobe. Kamacite occurs interstitial to silicates and troilite and as round globules, 30–110 μm is apparent diameter. Much of the kamacite has been replaced by limonite, a terrestrial weathering product that constitutes 28 wt% of the meteorite.

The high CaO and low MnO content of enstatite in Reckling Peak A80259 (Table 2) is similar to type 6 enstatite chondrites, although Abee (petrologic type 4) and St Sauveur (type 5) also contain enstatite with low MnO (6). Enstatite FeO content in Reckling A80259 is most similar to the two type 5 chondrites and is intermediate between that of type 4 and type 6 meteorites. The predominance of orthoenstatite and presence of ≥10% clinoenstatite is most similar to that of St Marks (type 5)[6]. (Type 6 chondrites contain almost entirely orthopyroxene and type 4 enstatite chondrites contain abundant clinopyroxene.) The kamacite Si content of Reckling Peak A80259 (2.1 wt%) is greater than that of type 6 chondrites (1.3 wt%) and less than that of type 4 or 5 chondrites (3.2 wt%)[6]. The presence of niningerite and absence of ferroan alabandite in Reckling Peak A80259 is a characteristic shared by the type 4 and 5 chondrites[6]. Some of the mineralogical characteristics of Reckling Peak A80259 are intermediate between type 4–5 and 6, but the absence of any large grains of plagioclase and the recrystallized, but readily delineated texture of the chondrules, unequivocally indicate type 5. Thus, the petrographic, mineralogical and crystallographic characteristics of Reckling Peak A80259 indicate that it is the third known type 5 enstatite chondrite.

The results of our determination of the bulk composition of the meteorite are presented in Fig. 1. In the same irradiation as that of Reckling Peak A80259 we analysed four samples of the Allende meteorite and found excellent agreement with literature values. Also shown in Fig. 1 are the mean compositions of the EH and EL groups, as calculated from the data compilations of Baedecker and Wasson[11] and Sears *et al.*[2], the compositions of the two previously known type 5 enstatite chondrites (St Marks and St Sauveur), and two other enstatite chondrites (Qingzhen and Allan Hills A77295) which were analysed at the same time and with the same standards and techniques as Reckling Peak A80259. Qingzhen and Allan Hills A77295 (paired with Allan Hills A77156) are two EH chondrites[12,13] whose analyses we publish here to demonstrate the absence of systematic errors in our data. The St Marks data are from various sources[2,11] and the St Sauveur data are from ref. 2. Siderophile, chalcophile and alkali elements have been plotted separately from refraction, lithophile and other elements because of their different abundance patterns in these meteorites. Within each division, elements are plotted in order of decreasing nebula condensation temperature. The element to Sc ratio for the sample has been divided by the same ratio for CI meteorites to obtain the value for the vertical axis. Normalization to a refractory element is the standard procedure for removing the effect of variations in major volatiles (for enstatite chondrites it is not

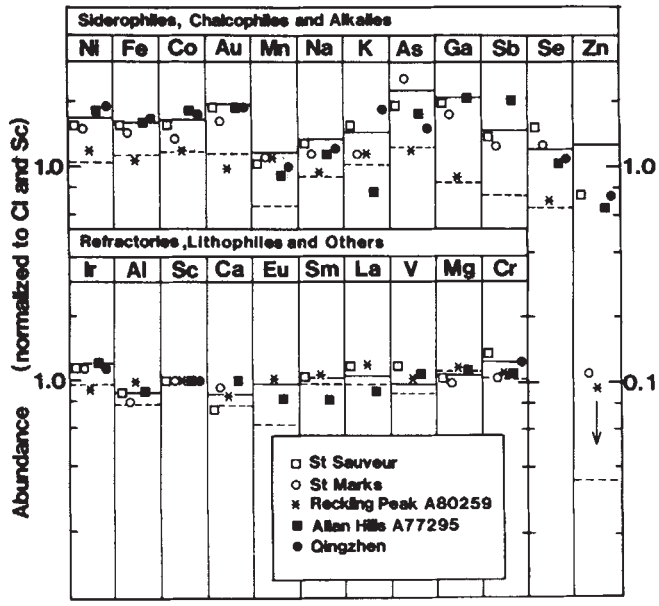


Fig. 1 Abundance pattern for selected elements in enstatite chondrites as determined by neutron activation analysis. The solid horizontal lines indicate mean values for the EH class and the broken horizontal values refer to mean values for the EL class.

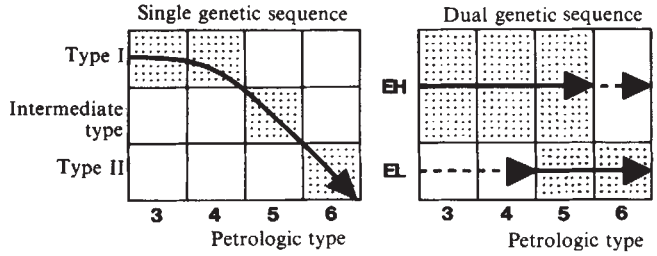


Fig. 2 Schematic diagram illustrating the two ways of viewing enstatite chondrite relationships. The existence of an EL5 is evidence for the dual genetic sequence.

normally an important normalization, but in the present case would remove any effects resulting from hydration during weathering; this is discussed below). The CI normalization removes abundance trends resulting from nucleosynthesis.

When normalized this way, EH chondrites are approximately 50% enriched in siderophile, chalcophile and alkali elements with respect to EL chondrites (Fig. 1). St Marks, St Sauveur, Qingzhen and Allan Hills A77295 usually plot close to the mean EH value, whereas Reckling Peak A80259 plots significantly below these meteorites, close to the mean EL value. In the one case where this does not happen (Mn), we suspect a problem with the literature value. Zinc is a special case, since it is a particularly volatile element and in ordinary chondrites its abundance varies systematically with petrologic type. This suggests that metamorphism may be involved in determining its abundance, rather than the factors (usually thought to be nebular processes) which result in the bimodal distribution of the other chalcophile elements.

Nickel, Co, Fe, Au and Ir are siderophile elements which lie within experimental error of the mean EL value. Iridium is unusual in that the difference between EH and EL is much less than for the other siderophiles—probably this is associated with its refractory nature—but nevertheless indicates an EL classification for Reckling Peak A80259. The chalcophiles—even Mn and the alkalies behave like chalcophiles in enstatite chondrites—are also consistent with an EL classification, with the exception mentioned above. Overall, the compositional data indicate that Reckling Peak A80259 is an EL chondrite.

A major potential problem in interpreting the present data is associated with the extensive weathering the meteorite has

© 1984 Nature Publishing Group

suffered. Biswas *et al.* discussed the effects of weathering on several Antarctic stone meteorites and concluded that any effects were minor[14]. However, their samples were less severely weathered than Reckling Peak A80259 (category 'A' compared with 'B/C' for Reckling Peak A80259 according to ref. 10). On the other hand, Bhandari *et al.*, when analysing the Parsa enstatite chondrite, argued that their data should be multiplied by 1.1 to compensate for weathering which had caused a dilution effect[15]. In fact, the concentrations we determined for Reckling Peak A80259 are 15–20% lower than mean EL for all elements; for example, our values for Fe, Co, Na and Sc are 218 mg g^{-1}, 668 μg g^{-1}, 4.60 mg g^{-1} and 5.93 μg g^{-1}, respectively, compared with EL group means of 265,780, 5.9 and 7.7, respectively. However, a dilution effect due to weathering will not affect Fig. 1 because of the Sc-normalization. It is possible, in principle, that a factor of 2 dilution could cause the siderophile elements in an EH chondrite to mimic EL values (we question that such extensive weathering would leave the rock recognizably meteoritic), but then the lithophile element concentrations would also be a factor of 2 lower than mean EH which is certainly not the case for Reckling Peak A80259. Interestingly, the mineral in enstatite chondrites that is most vulnerable to weathering is oldhamite (CaS) and, indeed, this mineral is absent from Reckling Peak A80259, yet the elements associated with oldhamite (the rare earth elements and much of the Ca) are present in normal EL abundances. Presumably, although decomposition of the oldhamite occurred, its decomposition products were not transported more than a few millimetres.

Figure 2 summarizes the relationship between enstatite chondrite classification and the question of whether there are one or two genetic sequences. The left-hand panel represents the single sequence model whereby type I → intermediate type → type II; the right-hand panel describes the two sequence models wherein EH3 → EH6 and EL3 → EL6. The occupied locations in the grid are shaded. The important point is that petrologic type 5 enstatite chondrites are now known to exist from both the EH and EL chemical groups. We suggest this is compelling evidence for two genetic sequences, analogous to the three sequences formed by H, L and LL ordinary chondrite cases. The observed tendencies for most EH chondrites to be of low petrologic type, and most EL chondrites to be of high petrologic type is also analogous to the situation for ordinary chondrites; for instance, most L chondrites (69%) are type 6, whereas most H chondrites (49%) are type 5[16].

We thank the meteorite working group of NASA/NSF and E. Scott, University of New Mexico, for supplying the meteorite samples used in this work, and D.W.G.S. thanks the Smithsonian Institution for hospitality. We also thank K. Keil and R. Hutchison for reviews. This work was supported by start-up funds from the University of Arkansas Chemistry Department and a grant from NASA (NAGW-296 to D.W.G.S.). The cost of irradiations is partially covered by US Department of Energy grant DEFG-0280-ER-10725 to the University of Missouri, research reactor facility.

Received 17 August; accepted 15 December 1983.

1. Van Schmus, W. R. & Wood, J. A. *Geochim. cosmochim. Acta* **31**, 747–765 (1967).
2. Sears, D. W., Kallemeyn, G. W. & Wasson, J. T. *Geochim. cosmochim. Acta* **46**, 597–608 (1982).
3. Yavnel', A. A. *Meteoritika* **23**, 32–41 (1963).
4. Anders, E. *Space Sci. Rev.* **3**, 583–714 (1964).
5. Mason, B. *Geochim. cosmochim. Acta* **30**, 23–39 (1966).
6. Keil, K. *J. geophys. Res.* **73**, 6945–6979 (1968).
7. Easton, A. J. *Meteoritics* **18**, 19–27 (1983).
8. Binz, C. M., Kurimoto, R. K. & Lipschutz, M. E. *Geochim. cosmochim. Acta* **38**, 1597–1606.
9. Dodd, R. T. *Meteoritics* (Cambridge University Press, 1981).
10. Score, R. A., Schwarz, C. M., Mason, B. & Bogard, D. D. *Antarctic Meteorite Newsl.* **5**(1), (1982).
11. Baedecker, P. A. & Wasson, J. T. *Geochim. cosmochim. Acta* **89**, 735–765 (1975).
12. Rambaldi, E. R., Rajan, R. S. & Wang, D. *Lunar planet. Sci.* **14**, 626–627 (1983).
13. McKinley, S. G., Keil, K. & Scott, E. R. D. *Meteoritics* **17**, 251 (1982).
14. Biswas, S., Ngo, H. T. & Lipschutz, M. E. *Z. Naturforsch.* **35a**, 191–196 (1980).
15. Bhandari, N., Shah, V. B. & Wasson, J. T. *Meteoritics* **15**, 225–233 (1980).
16. Motylewski, K. *Cambridge Chondrite Compendium* 2nd edn (Smithsonian Astrophysical Observatory, Cambridge, Massachusetts, 1978).

Chondrite thermal histories constrained by experimental annealing of Quenggouk orthopyroxene

J. R. Ashworth*, L. G. Mallinson*, R. Hutchison† & G. M. Biggar‡

* Department of Geological Sciences, University of Aston, Gosta Green, Birmingham B4 7ET, UK
† Department of Mineralogy, British Museum (Natural History), Cromwell Road, London SW7 5BD, UK
‡ Grant Institute of Geology, University of Edinburgh, West Mains Road, Edinburgh EH9 3JW, UK

The meteorites called 'ordinary chondrites' preserve a record of processes in the formation of their parent bodies (asteroids or similar objects) in the early Solar System. An important aspect yet to be explained is a series of 'petrologic types'[1], numbered 3 to 6, with increasingly homogeneous minerals and annealed textures. The annealing (which postdates the highest-temperature events that produced chondrules as droplets of melt) is usually regarded as a prograde metamorphism, that is, reheating after cold accretion of the parent bodies[2], but evidence has been presented for origin of the textures and mineral compositions during primary cooling (see, for example, ref. 3). Here we report the first constraint on chondrite 'metamorphic' history from experimental annealing of a silicate mineral in natural meteoritic specimens. 'Striated' orthopyroxene was converted almost entirely to normal orthopyroxene in 1 week at approximately 800 °C, without detectable change of composition. This result tends to support the primary cooling hypothesis.

Orthopyroxene is essentially a solid solution between $MgSiO_3$ (En) and $FeSiO_3$ (Fs) with minor $CaSiO_3$ (Wo). In chondrites of intermediate petrologic types (4 and 5) it is often optically striated[4] between crossed polars, different striae extinguishing in slightly different positions. Transmission electron microscopy (TEM) has shown that the 'striation' is due to lamellar 'defects' parallel to (100). Although not fully resolved by conventional TEM[5], individual defects are interpreted as lamellae having the clinopyroxene structure, in two (twinned) orientations sharing (100) planes with the host orthopyroxene[5]. Such alternation of ortho and clino polymorphs on the unit-cell scale has been resolved by high-resolution TEM[6] in comparable material prepared from terrestrial orthopyroxene by heating then rapidly cooling.

Striated orthopyroxene is interpreted as an inversion product of protopyroxene which crystallized from the melt in droplet chondrules[5]. Protopyroxene inverts to normal orthopyroxene if cooled slowly. At the other extreme, on very rapid quenching, it inverts entirely to twinned clinopyroxene. The experimental analogue of striated orthopyroxene, characterized by X-ray diffraction as a mixture of ortho and clino, is formed at cooling rates <1 °C per h (ref. 7) (through an inversion temperature which is not well known for Fs-bearing compositions but is ~1,000 °C in pure En)[7]. In previous annealing experiments of short duration (hours to days) near 1,000 °C, twinned clinopyroxene has been converted to the stable form, orthopyroxene[8,9], and lamellae of clinopyroxene (formed by deformation) in orthopyroxene have been partly removed[10,11]. Thus, striated orthopyroxene is expected to react towards normal orthopyroxene if annealed sufficiently.

We sought the effect at relatively low temperatures approximating the accretion temperature of 800 °C which has been inferred[12] for one chondrite of the H chemical group[1]. Quenggouk (British Museum (Natural History) specimen 33764), another H-group chondrite (petrologic type 4), was selected for the experiments because its chondrule orthopyroxene is well striated and has been described in detail[5]. Like many H-group chondrites, it lacks complicating effects of shock-heating[13]. Aliquots (~100 mg) were sealed in Pt capsules,

© 1984 Nature Publishing Group

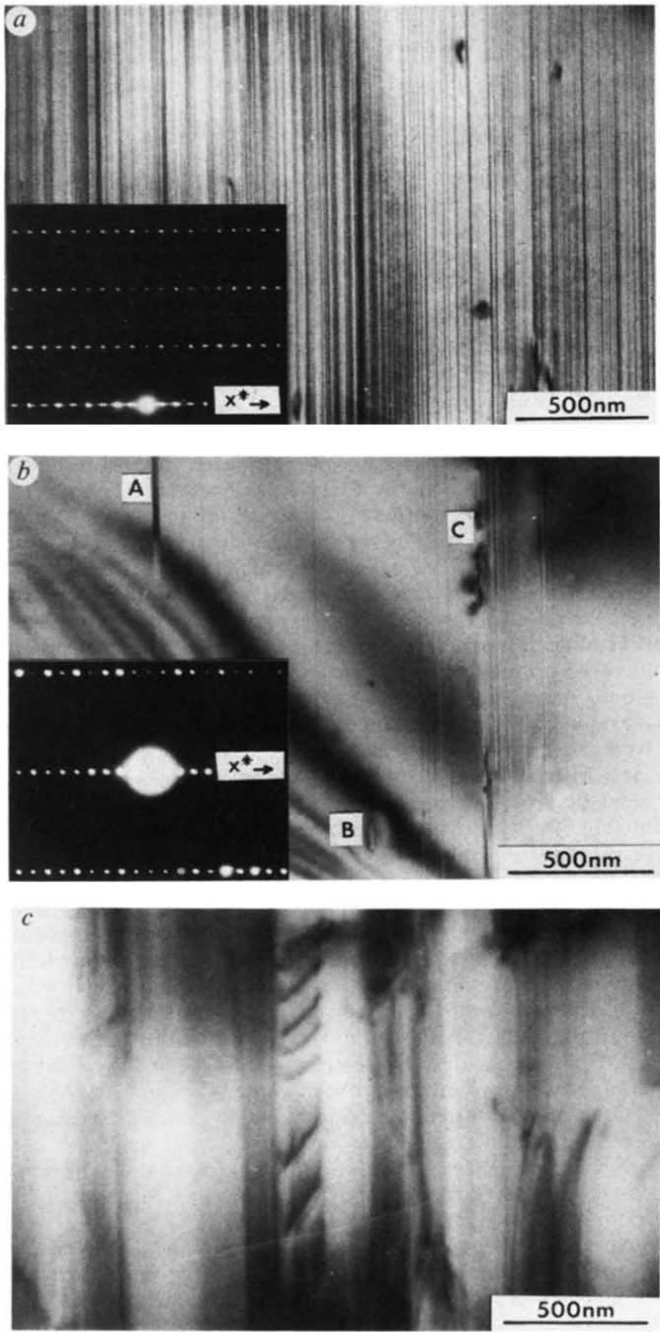


Fig. 1 High-voltage TEM photomicrographs of experimentally annealed Quenggouk orthopyroxene viewed under general diffraction conditions. *a*, Unmodified striated structure in the specimen annealed at 695 °C for 4 weeks. Inset [011] diffraction pattern shows diffuse maxima elongated parallel to x^*. *b*, Predominantly normal orthnopyroxene (inset, [010] diffraction pattern) with a few clino lamellae, in the specimen annealed at 790–800 °C for 1 week. Note thick lamella at A, isolated unit dislocation at B and strain contrast near lamella boundary C due to partial dislocations in the boundary. *c*, Oblique view of the same specimen showing partial dislocations in the boundary of the central lamella.

surrounded by Fe powder inside a second Pt capsule at about 15 bars (argon cylinder pressure) to prevent bursting during the annealing runs. One run was at 695 °C for 4 weeks, the other at 790–800 °C for 1 week. The friable specimens were then impregnated with epoxy resin, thin-sectioned and examined optically before ion-thinning for TEM. High-voltage (1 MV) TEM was used, preserving exact comparability with the previous study of untreated Quenngouk[5]. Additionally, energy-dispersive X-ray microanalysis was performed in a different TEM at 100 kV.

One striated-orthopyroxene chondrule was prepared for TEM from each experimental run. The orthopyroxene grains have the usual size (~100 μm) and tabular shapes. In the 695 °C specimen, the optical striation appeared normal for Quenggouk, but in the 790–800 °C material it was faint, with very slight variations in extinction.

TEM shows that the 695 °C experiment produced a null result: the striated orthopyroxene is indistinguishable from that in natural Quenggouk[5]. Figure 1*a* shows the pervasiveness of the (100) lamellar contrast, associated with diffuseness in electron diffraction patterns (Fig. 1*a*) as previously documented in untreated Quenggouk[5]. The product of the higher-temperature experiment is very different. It has properties not found in natural Quenggouk[5], but analogous to annealing effects in orthopyroxene where clino lamellae had been introduced by deformation[11], and therefore also expected where striated orthopyroxene has reacted towards normal (defect-free) orthopyroxene. Regions of perfect orthopyroxene giving a sharp diffraction pattern are interrupted by batches of lamellae which are often rather thick (Fig. 1*b*). Viewed obliquely, individual lamellae are resolved (Fig. 1*c*) and the thicker ones are seen to have linear features in their walls (Fig. 1*c*) which, by comparison with published studies[5,11], are clearly partial dislocations, marking unit-cell steps of thickness of the lamella. These properties were found throughout several large orthopyroxene grains.

Analysis in the TEM gives the composition $En_{85}Fs_{14}Wo_1$ in the 695 °C specimen and $En_{83}Fs_{16}Wo_1$ in the 790–800 °C specimen, the same (within analytical uncertainty) as previously found[5] by electron microprobe analysis of striated orthopyroxene in untreated Quenggouk. No compositional variations were detected at the lamellae: even the thick lamellae (Fig. 1*b*, *c*) are clinopyroxene of approximately the same composition as the host orthopyroxene, not exsolution lamellae of any sort. Thus, the experiment produced no detectable compositional changes.

Persistence of a few rather thick lamellae is also characteristic of orthopyroxene annealed after deformational production of clino lamellae[5,11] but in this case numerous isolated (unit) dislocations are also found. This is attributable to the locally untwinned nature of the clino component produced by shearing[11] or shock[5]; annealing of the intimately twinned analogue in striated orthopyroxene will tend to generate unit dislocations of opposite sign, but these can mutually annihilate, leaving few such dislocations in the end product (Fig. 1*b*).

The experiments show that Quenggouk has not been naturally held at 800 °C for as long as 1 week. Such annealing would have converted its striated orthopyroxene to a product structurally indistinguishable from the orthopyroxene in higher petrologic types (5 to 6).

The temperature constraint is superficially consistent with prograde metamorphic temperatures favoured by Dodd[2], but these are based on pyroxene compositions, in particular the Ca content of orthopyroxenes (supposedly at equilibrium with the coexisting calcic pyroxenes) in LL-group chondrites[14]. Now, in the H group at least, Ca contents of coexisting pyroxenes[3,5] are not consistent with the equilibrium phase diagram at metamorphic temperatures[15] and are interpreted as being inherited from the time when the grains grew from chondrule melts[3]. The present work shows how striated and non-striated orthopyroxenes can inherit the same compositions. (The slight differences in annealing history can occur during continuous cooling[7]; re-heating is unnecessary.)

The lack of compositional change in the experiments is consistent with diffusion data[16,17] which (extrapolated down to 800 °C) indicate a time scale many orders of magnitude greater than 1 week for Ca diffusion through ~100 μm. We can tentatively extend this comparison of time scales to lower temperatures, since the temperature dependence of the time scale for clino-lamella removal can be estimated by combining our data with higher-temperature studies on pyroxenes of approximately the same composition[10,11]. Although few data are avail-

© 1984 Nature Publishing Group

able, this combination gives a rough estimate of 180–280 kJ mol^{-1} for the activation energy for clino-lamella removal, which is comparable with or slightly less than activation energies for bulk diffusion[16,17]. If confirmed by further experiments, this result will indicate that the time scale contrast persists throughout the metamorphic temperature range, so orthopyroxenes could not be metamorphically homogenized without losing their striated structure. Thus, we suggest that the series of increasing pyroxene homogeneity with increasing petrologic type[1] is unlikely to be due to prograde metamorphism, since the orthopyroxenes often retain striated structure[4]. On the other hand, both the compositional and structural variations in the orthopyroxenes can be explained by crystallization at varying degrees of departure from melt-crystal equilibrium, followed by variable degrees of annealing during cooling through a temperature range approximated in our experiments.

This interpretation also fits the data on metallic minerals in chondrites. Bevan and Axon[18] argued that metal grains in H-group chondrites are solidified melts of variable composition on which were superimposed the effects of solid state diffusion. A recent experimental study[19] showed that for cooling rates <1,000 °C Myr^{-1}, metal grains homogenize. Our new evidence suggests a much faster cooling rate and supports the conclusions of Bevan and Axon[18].

This work was supported by NERC. TEM facilities were provided by Professor M. H. Loretto.

Received 14 October 1983; accepted 26 January 1984.

1. Van Schmus, W. R. & Wood, J. A. *Geochim. cosmochim. Acta* **31**, 747–765 (1967).
2. Dodd, R. T. *Meteorites. A Petrologic-Chemical Synthesis* (Cambridge University Press, 1981).
3. Hutchison, R., Bevan, A. W. R., Agrell, S. O. & Ashworth, J. R. *Nature* **287**, 787–790 (1980).
4. Binns, R. A. *Miner. Mag.* **37**, 649–669 (1970).
5. Ashworth, J. R. *Earth planet. Sci. Lett.* **46**, 167–177 (1980).
6. Iijima, S. & Buseck, P. R. *Am. Miner.* **60**, 758–770 (1975).
7. Smyth, J. R. *Am. Miner.* **59**, 345–352 (1974).
8. Sadanaga, R., Okamura, F. P. & Takeda, H. *Miner. J.* **6**, 110–130 (1969).
9. Buseck, P. R. & Iijima, S. *Am. Miner.* **60**, 771–784 (1975).
10. Boland, J. N., Duba, A. & Eggleton, A. *J. Geol.* **82**, 507–514 (1974).
11. Coe, R. S. & Kirby, S. H. *Contr. Miner. Petrol.* **52**, 29–55 (1975).
12. Hutchison, R., Bevan, A. W. R., Agrell, S. O. & Ashworth, J. R. *Nature* **280**, 116–119 (1979).
13. Ashworth, J. R. *Proc. R. Soc.* A**374**, 179–194 (1981).
14. Heyse, J. V. *Earth planet. Sci. Lett* **40**, 365–381 (1978).
15. Lindsley, D. H. *Am. Miner.* **68**, 477–492 (1983).
16. Freer, R., Carpenter, M. A., Long, J. V. P. & Reed, S. J. B. *Earth planet. Sci. Lett.* **58**, 285–292 (1982).
17. Brady, J. B. & McCallister, R. H. *Am. Miner.* **68**, 95–105 (1983).
18. Bevan, A. W. R. & Axon, H. J. *Earth planet. Sci. Lett.* **47**, 353–360 (1980).
19. Willis, J. & Goldstein, J. I. *Nature* **293**, 126–127 (1981).

Molecular fractal surfaces

David Avnir & Dina Farin

Department of Organic Chemistry, The Hebrew University of Jerusalem, Jerusalem 91904, Israel

Peter Pfeifer

Fakultät für Chemie, Universität Bielefeld, D-4800 Bielefeld, FRG

In this letter we report that, at the molecular-size range, the surfaces of most materials are fractals, that is, at this range, surface geometric irregularities and defects are characteristically self-similar upon variations of resolution. The whole range of fractal dimension[1], $2 \leq D < 3$, is found in the many examples presented. Two representative examples, namely adsorption of polystyrene on alumina and adsorption of krypton on dolomite are discussed in some detail. Our findings suggest a simple solution to the problem of quantifying the degree of surface irregularity[2,3] at a resolution which is of relevance to many aspects of surface science. The results provide a general explanation for phenomenological links between various surface parameters, and we derive a set of equations of use in predicting surface variables.

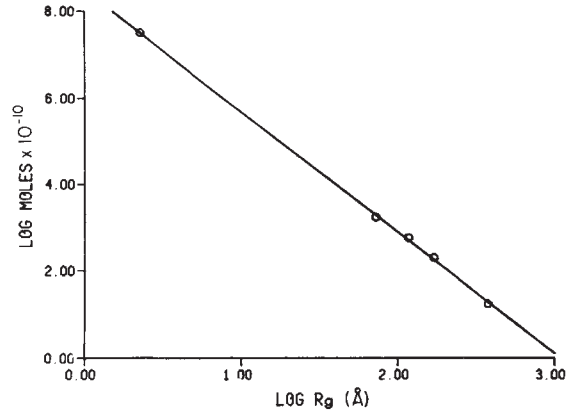


Fig. 1 Aluminium oxide, $D = 2.79 \pm 0.03$. Monolayer amount of polystyrene molecules and nitrogen (mol $\times 10^{-10}$ per g), as a function of the radius of gyration R_g (Å).

Surface-science has in the past tended to deal with the problem of description of geometrically irregular surfaces, such as those found in oxide gels, carbons, catalysts, metals[4], by treating irregularity as a deviation from well-defined Eucleidean shapes. An entirely different approach is possible, if the surface is (statistically) invariant over a certain range of scale transformations—that is, if the surface is self-similar upon changes in resolution power. For such surfaces, the degree of irregularity can be given by one number: the fractal dimension[1], $2 \leq D < 3$, of the surface. The higher the D value, the more wiggly and space filling the surface is (see below). The obvious question is then, is self-similarity common in nature so as to make the fractal-approach useful? The similarity of computer-generated fractal-objects to real objects led Mandelbrot[1] to advocate strongly a positive answer to that question; and in recent years several reports have appeared[1,5–7] in which boundary lines of natural objects ('coast lines') have been found to be fractals ($1 \leq D < 2$). Similarly, natural landscapes were reported to be fractals[8,9]. On the molecular scale, the only direct experimental fractal measurement is, to our best knowledge, the analysis of back-bone 'wiggliness' in proteins[10].

Here we present an extensive compilation suggesting that the fractal approach is indeed of value. Moreover, in contrast to most previous fractal analyses, which probed the $1 \leq D < 2$ range of fractal dimensions, and could only indirectly provide the surface fractal dimensions, our methods directly probe the $2 \leq D < 3$ range.

Table 1 Equations derived for determination of fractal dimension

Equation no.		Comments
(1)	$n \propto r^{-D}$	See text
(2)	$n \propto \sigma^{-D/2}$	—
(3)	$A \propto \sigma^{(2-D)/2}$	—
(4)	$n \propto R_g^{-D}$	—
(5)	$n \propto v^{-D+1}$	Holds only for series of homologous linear absorbates lying flat on surface
(6)	$V_m \propto \sigma^{-D/2}$	—
(7)	$A \propto R^{D-3}$	—
(8)	$n \propto R^{D-3}$	—
(9)	$\sigma_0 \leq \sigma \leq \sigma_0 (R_{max}/R_{min})^2$	—

In these equations, n is the monolayer coverage (for example, moles of adsorbate per gram adsorbent; σ is the effective cross-sectional area of adsorbate; A is the apparent surface area ($m^2\ g^{-1}$); R_g is the radius of gyration; v is the molar volume; V_m is the monolayer coverage as expressed by the STP volume in the BET method; R, R_{max} and R_{min} are values for particle radius; σ_0 is the cross-sectional area of adsorbate, as used for equations (7) and (8).

© 1984 Nature Publishing Group

Table 2 Fractal dimensions in a range of materials

Fractal dimension*	Found in	At the range (Å²)†	Adsorbates	Method of D determination‡	Ref.§
High					
2.91±0.02	Upper Columbus dolomitic rock from Bellevue, Ohio	20–47,000	Kr	(7)	13
2.97±0.01	Goodland high calcium rock, from Idabel, Okla.	20–47,000	Kr	(7)	13
2.88±0.02	Granitic rock from SHOAL nuclear test site, Nevada	16–16,500	N_2	(7)	14
2.73±0.05	Igneous rock sample from SHOAL site	14–14,300	Ar	(7)	14
2.71±0.14	Granular activated carbon-Tsurumi HC-8, from coconut shell	(16–37)	N_2, 12 various organic molecules	(2)	15
2.80±0.16	Granular activated carbon—Fujisawa B-CG, from coconut shell	(16–37)	N_2, 12 various organic molecules	(2)	15
2.90±0.01	Carbonate rock from groundwater test well, Yucca Flat, Nevada	16–16,500	N_2	(7)	14
2.92±0.02	Soil (kaolinite, trace halloysite)	150–16,500	Malachite green	(8)	16
2.94±0.04	Porous silicic acid	(16–34)	N_2, MeOH, EtOH, PrOH, iso-PrOH, BuOH	(3)	2
2.79±0.03	Activated alumina grade F-20 (Alcoa Corp.)	16–45,100	Polystyrene fractions, N_2	(4)	12
2.78±0.21	Charcoal (BDH) of animal origin	1,400–180,000	Styrene-methyl-methacrylate random copolymers	(4)	11
2.67±0.16	Porous coconut charcoal (Standard Chemical Co., Montreal)	(16–47)	N_2, C_2H_2, C_2H_4, CH_4, C_2H_6, C_3H_8, n-C_4H_{10}, iso-C_4H_4	(2)	11
Medium					
2.57±0.04	Porous α-FeOOH pigment for magnetic tapes	16–980	N_2	(7)	17
2.35±0.11	Crushed Corning 0010 lead glass	21–14,900	Hexadecyl-1-^{14}C-trimethyl ammonium bromide	(7)	11
2.52±0.07	Coal mine dust from Western Pennsylvania	16–180	N_2	(7)	18
2.33±0.08	Coal mine dust from Western Pennsylvania	16–270	N_2	(7)	18
2.25±0.09	Carbon black	(16–71)	N_2, benzene, naphthalene, anthracene, phenantrene	(2)	11
2.54±0.12	Slightly porous coconut charcoal (Standard Chemical Co., Montreal)	(16–47)	N_2, C_2H_2, C_2H_4, CH_4, C_2H_6, C_3H_8, C_4H_{10}, iso-C_4H_{10}	(2)	11
2.30±0.07	Slightly porous coconut charcoal (Standard Chemical Co., Montreal)	(16–47)	N_2, C_2H_2, C_2H_4, CH_4, C_2H_6, C_3H_8, C_4H_{10}, iso-C_4H_{10}	(2)	11
2.63±0.03	Mosheim high calcium from Stephens City, Va.	20–47,000	Kr	(7)	13
2.58±0.01	Niagara (Guelph) dolomite from Woodville, Ohio	20–47,000	Kr	(7)	13
2.46±0.11	Glassy melted rock from Rainier nuclear zone, Nevada	14–14,300	Ar	(7)	14
2.29±0.06	Soil (mainly feldspar quartz and limonite)	150–21,800	Malachite green	(8)	16
Low					
2.02±0.06	Aerosil—nonporous fumed silica (Degussa)	16–529	N_2	(7)	11
2.15±0.06	Snowit—ground fine Belgian quartz glass of high purity	16–10,600	N_2	(7)	19
2.14±0.06	Madagascar quartz from Thermal Syndicate	16–1,850	N_2	(7)	19
1.95±0.04	Periclase—electrically fused and crushed magnesite	16–720	N_2	(7)	20
2.02±0.05	Synthetic faujasite ($Na_2O{\cdot}Al_2O_3{\cdot}2{,}67SiO_2{\cdot}mH_2O$) (Linde Air Products)	(16–68)	N_2, n-alkanes (C_1, C_{3-7})	(5)	11
2.07±0.01	Graphite—Vulcan 3G (2700) (National Physical Lab., Teddington, UK)	(16–178)	N_2, n-alkanes (C_{22}, C_{28}, C_{32})	(2)	11
2.04±0.16	Graphon—partially graphitized carbon black formed by heating MPC black to 3,200 °C	(15–41)	N_2, MeOH, EtCl, benzene, cyclohexane	(6)	11
2.13±0.16	Graphon—graphitized carbon black (Cabot Corp.)	1,400–180,000	Styrene-methyl-methacrylate random copolymers	(4)	11
2.04±0.04	Active, (nonporous), coconut charcoal (Standard Chemical Co., Montreal)	(16–47)	N_2, C_2H_2, C_2H_4, CH_4, C_2H_6, C_3H_8, n-C_4H_{10}, iso-C_4H_{10}	(2)	11
1.97±0.02	Active, (nonporous), coconut charcoal (Standard Chemical Co., Montreal)	(16–47)	N_2, C_2H_2, C_2H_4, CH_4, C_2H_6, C_3H_8, n-C_4H_{10}, iso-C_4H_{10}	(2)	11
2.16±0.04	Iceland spar, massive calcite from Chihuahua, Mexico	20–47,000	Kr	(7)	13

* Error estimation indicates the least-square analysis of best fit to a straight line.

† Values in parentheses give the range of yardstick sizes, when the full range cannot be estimated directly either from equation (9) or from polymer adsorption. For nitrogen, the standard used value of 16.2 Å² is taken.

‡ Numbers refer to equation nos in Table 1.

§ All experimental data have been published previously: refs 2, 11 contain details on data and its analyses; refs 12–20 are data sources only.

The fractal theory of surface science we have developed[2,3] is an extension of the basic relation[1] which describes self-similarity:

$$n \propto r^{-D} \tag{1}$$

where n is the number of yardsticks of size r necessary to cover (in order to measure) the length or the surface area of the fractal object. Note that for, say, flat smooth surfaces, equation (1) becomes the familiar $n \propto r^{-2}$, and that volume-filling is $n \propto r^{-3}$. In our application, the yardsticks are molecules. The equations used are shown in Table 1.

Following our finding that silica gel has a fractal surface[2], we found the surfaces of some other common materials to be fractals (ref 11 and Table 2). Further analysis of published data (Table 2) lead us to conclude that at the molecular range, the surfaces of most materials are fractals. Well over 75% of the analysed surfaces fell into this category. Table 2 emphasizes the wide diversity of materials which possess a fractal surface, and the ability of our method to probe the irregularity of hindered areas within porous materials. The complete range ($2 \leq D < 3$) of fractal dimensions seems to be found in nature: low $D = 2.0$ values, indicate regularity and smoothness (for example, Aero-

© 1984 Nature Publishing Group

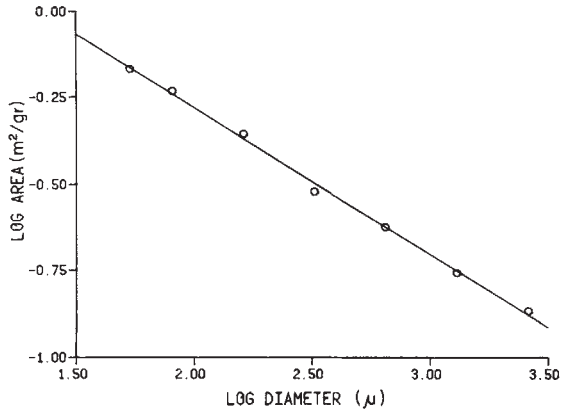


Fig. 2 Niagara dolomite, $D = 2.58 \pm 0.01$. BET (Kr) surface area (m^2 per g) shown as a function of particle diameter (μm).

sil), D values deviating slightly from 2 are indicative of a surface with slight defects (2.07 for graphite), intermediate D values represent a fairly irregular surface (some types of charcoal), and D values close to 3 indicate extreme irregularity of the surface (silica gel). A wide range of self similarity can be detected when equations (7) or (8) are used, or when the yardsticks are polymer fractions. The smaller ranges found in Table 2 are artificially limited by the given range of yardstick sizes. The processes by which these fractal surfaces were formed probably operated over a much wider scale, as can be judged by comparison of ranges of similar materials, such as charcoals. In any event, even the limited-range examples carry information of interest, since they cover the size of all small organic molecules (C_1 to C_{10}).

Two examples taken from Table 2 serve to illustrate the nature of those fractal surfaces. The first is the surface of porous alumina, (Alcoa grade F-20) studied by Burns and Carpenter[12]. They adsorbed monolayers of nitrogen and four spherically coiled polystyrene fractions, having radii of gyration in the range 73–379 Å. The decrease in the resulting apparent surface area, as a function of yardstick size, is faster than $n \propto r^{-2}$ (for smooth surfaces), and obeys equation (4), as depicted in Fig. 1. From the slope of the straight line, the fractal dimension of the surface of the studied alumina is 2.79 ± 0.03. The range of surface self-similarity found for this example is high, covering three orders of magnitudes. This observation was made possible by the remarkable experimental result that monolayer values obtained by different methods (Brunauer, Emmett and Teller (BET) surface area test at cryogenic temperatures for the small nitrogen; adsorption from cyclohexane solution for the large polymer yardsticks) fit closely the same straight line—that is, the geometric factor rather than the nature of physisorption, dominates. The self-similarity of this surface is further demonstrated by the fact that whereas the large polymers can penetrate and probe only the widest pores (and the outer surface), nitrogen probes much of the pore network that is hidden from the polymer.

The second example involved a different method of D-determination, the analysis of change in apparent surface area as a function of particle size[3] (equations (7) and (8)). The method has several advantages: (1) it is easier to obtain a wide range of particle sizes than to use a wide range of molecular yardsticks, so that, a wide range of self-similarity may be detected; (2) adsorption of only one type of molecule is necessary; (3), the measurement scale in this approach is the particle-radius so that an exact knowledge of the effective cross section of the probe molecule is necessary only for estimation of the range of self-similarity (equation (9)). Figure 2 summarizes the example, in which Love and Whittaker measured the surface area of limestones from various sources by krypton adsorption[13]. Niagara (Guelph) dolomite from Woodville, Ohio, was crushed and sieved into seven fractions ranging from 53.5 to 2,605 μm in size, and from equation (7) we find $D = 2.58 \pm 0.01$. All the other rock-surfaces studied by Love and Whittaker were found to be fractals (see all Kr entries in Table 2).

Thus the geometry of materials has conventionally been classified in terms of types of crystallinity or amorphousness; we now propose fractality as a new viewpoint in material geometry classification. Since it is difficult to conceive of any dynamic interfacial process that is independent of surface irregularity or surface defects, the importance of such classification seems obvious. Catalysis is one phenomenon which is governed by such surface characteristic and this is now being studied.

This work was supported by the Fritz Haber Research Center for Molecular Dynamics, Jerusalem, the Israel–US Binational Foundation (D.A.) and by the Ben-Gurion Foundation (to D.F.). D.A. is an M. Richter Fellow.

Received 23 August; accepted 21 December 1983.

1. Mandelbrot, B. B. *The Fractal Geometry of Nature* (Freeman, San Francisco, 1982).
2. Avnir, D. & Pfeifer, P. *Nouv. J. Chim.* **7**, 71–72 (1983).
3. Pfeifer, P. & Avnir, D. *J. chem. Phys.* **79**, 3558–3565 (1983).
4. Beddow, J. K. & Meloy, T. P. (eds) *Advanced Particulate Morphology* (CRC, Florida, 1977).
5. Lovejoy, S. *Science* **216**, 185–187 (1982).
6. Paumgartner, D., Losa, G. & Weibel, E. R., *J. Microsc.* **121**, 51–63 (1981).
7. Kaye, B. H. *Direct Characterization of Fine Particles* (Wiley, New York, 1981).
8. Berry, M. V. & Hannay, J. H. *Nature* **273**, 573 (1978).
9. Burrough, P. A. *Nature* **294**, 240–242 (1981).
10. Allen, J. P., Calvin, J. T., Stinson, D. G., Flynn, C. P. & Stapelton, H. J. *Biophys. J.* **38**, 299–310 (1982).
11. Avnir, D., Farin, D. & Pfeifer, P. *J. chem. Phys.* **79**, 3566–3571 (1983).
12. Burns, H. Jr & Carpenter, D. K. *Macromolecules* **1**, 384–390 (1968).
13. Love, K. S. & Whittaker, C. W. *Agr. Food. Chem.* **2**, 1268–1272 (1954).
14. Carnahan, C. L., Castagnola, D. C. & Smith, M. E. U.S. Atomic Energy Comm. NVO-1229-98 (1968).
15. Urano, K., Omori, S. & Yamamoto, E. *Env. Sci. Tech.* **16**, 10–14 (1982).
16. Cohen, W. H. & Knight, B. H. *J. Soc. Chem. Ind.* **66**, 357–364 (1947).
17. Suzuki, S. *et al. Ferrites*, 556–560 (Tokyo, 1980).
18. Corn, M. *et al. Am. ind. Hygiene J.* **34**, 279–285 (1973).
19. Cartwright, J., Wheatley, K. & Sing, K. S. W. *J. appl. Chem.* **8**, 259–264 (1958).
20. Girgis, B. S. & Girgis, L. G. *Powder Tech.* **7**, 85–91 (1973).

A sampling artefact affecting the uranium content of deep-sea porewaters obtained from cores

J. Toole*, J. Thomson†, T. R. S. Wilson† & M. S. Baxter*

* Department of Chemistry, The University, Glasgow G12 8QQ, UK
† Institute of Oceanographic Sciences, Brook Road, Wormley, Godalming, Surrey GU8 5UB, UK

A wide range of uranium concentrations (1.3–65 μg l^{-1}) have been reported in deep-sea porewaters extracted from cores[1–5], values which contrast with the accepted constant figure of 3.3 ± 0.2 μg l^{-1} for the uranium content of open ocean seawater[4,6–8]. In seawater the constant level of uranium is believed to be due to the conservative behaviour of the anionic uranyl carbonate complex[9,10]. In a recent review Cochran[11] concludes that porewater uranium contents higher than the seawater value are found in association with an anoxic condition in the sediments, while most oxic sediments have lower values in the range 2–2.5 μg l^{-1}. This latter observation caused Cochran and Krishnaswami[5] to suggest that uranium uptake by some sediment solid phase must occur and ought to be included in the oceanic uranium balance[4,7]. We present here detailed uranium porewater profiles for two contrasting eastern Atlantic sediments from an *in situ* porewater sampler and from squeezed corer sediment. These indicate that lower values in oxic sediments are almost certainly caused by a sampling artefact related to core decompression.

The porewaters were collected on RRS *Discovery* Cruise 129 in 1982, using the Institute of Oceanographic Sciences (IOS) Mark II *in situ* porewater sampler[12], a box corer[13] to retrieve short sediment sections with undisturbed interfaces, and a Kastenlot corer to recover longer (2 m) sections. Subcores were taken as soon as possible after retrieval using core liner, extruded

© 1984 Nature Publishing Group

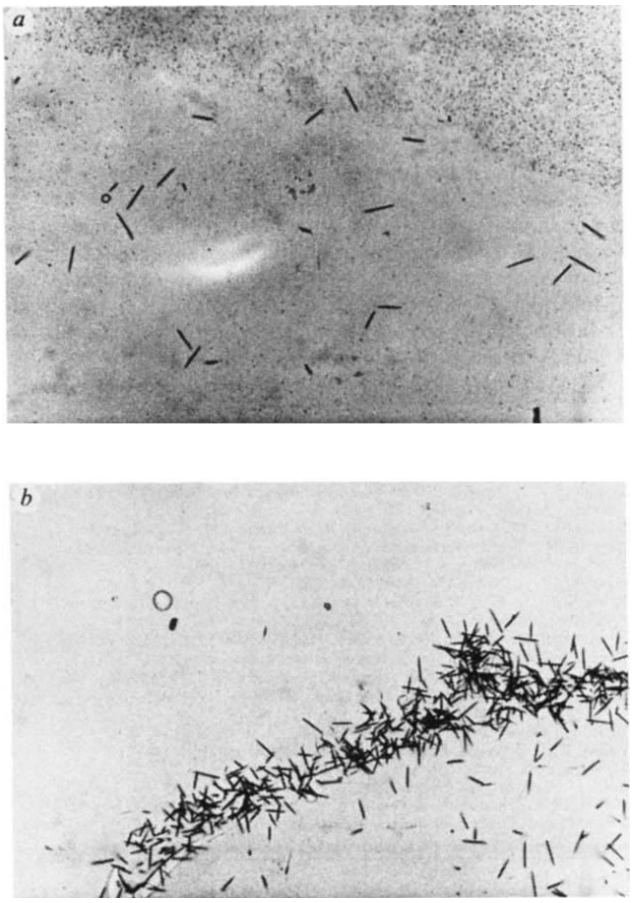


Fig. 1 *a*, Field of view showing detector track distribution at the edge of a seasalt deposit with added ethylene glycol (×200). Uranium concentration 3.3 $\mu g\,l^{-1}$. *b*, Field of view showing detector track distribution at the edge of a seasalt deposit without spreading agent (×200). Uranium concentration 3.3 $\mu g\,l^{-1}$.

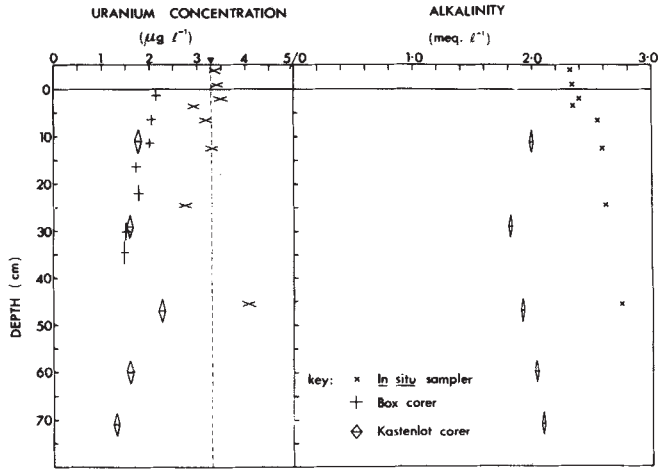


Fig. 2 Porewater uranium and alkalinity profiles for station 10552, Cape Verde Abyssal Plain (see Table 1 for individual positions).

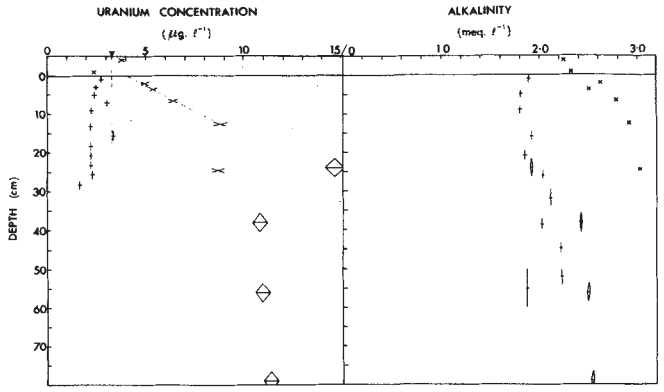


Fig. 3 Porewater uranium and alkalinity profiles for station 10554, Madeira Abyssal Plain east of Great Meteor Seamount. (Key as Fig. 2: see Table 1 for individual positions.)

in a glove box and transferred to squeezer units under a nitrogen atmosphere, then held at bottom water temperature (4 °C) until hydraulic squeezing was complete. Both *in situ* and squeezer filtrations were performed using 0.4 μm Nuclepore filters.

Uranium determinations were made by the sensitive fission track method using Lexan polycarbonate detectors and an adaptation[14] of Fleischer's method[15,16] for seawater. Fifty fields of view were counted across samples which had been etched after exposure to a thermal neutron (n) dose of 2×10^{16} n cm^{-2}. Track counting was either by optical microscope (total area 8.309 mm^2) or by Quantimet image analyser (total area 0.699 mm^2), depending on track density. A total of about 1,300 tracks were counted for a uranium concentration of 3.3 $\mu g\,l^{-1}$, yielding a replicate error of 6.1%. Figure 1*a* shows the track distribution at the edge of a deposit in which glycol spreading agent had been used to prevent the formation of an outer rim of salt crystals during evaporation of the sample. In this case a uniform distribution of tracks is obtained, simplifying analysis to one of comparative track densities between samples and standards instead of total track counting. Without the glycol additive, uranium concentrates on the outer rim (Fig. 1*b*). Since track clustering was not observed in this study, it may be concluded that colloidal aggregates of uranium were absent.

At station 10552 on the Cape Verde Abyssal Plain the sediments consist of a calcareous marl/ooze (36–78% $CaCO_3$) with a low mean sediment accumulation rate (0.4 cm per 10^3 yr determined from $^{230}Th_{excess}$ data). Oxygen measurements on the porewaters[17] show that free oxygen is present, falling from 66% of bottom water values just below the sediment surface to 43% at the deepest point sampled (1.7 m). Extrapolation of the profile trend suggests that the sediment column may be oxic to at least 10 m. There is good agreement between the levels of the nutrients silicate and phosphate for the box core, Kastenlot core and porewater sampler at this station. Marked discrepancies in alkalinity and in uranium content are observed, however, between the profiles from the corers and that from the *in situ* sampler (Fig. 2, Table 1). The loss of alkalinity from corer porewaters during recovery is now well documented[18–22] and is believed to be due to $CaCO_3$ precipitation during the pressure change on core retrieval from the sea floor. Initial attempts to correct for the pressure effect by a thermodynamic treatment[18,20] have been shown to be not widely applicable[21,22] perhaps because equilibrium is not attained before sediment squeezing. In the present case, the uranium loss is relatively larger than the alkalinity loss. The core samples are reduced by a mean 44% in uranium content, and by 27% in alkalinity, relative to the pore water sampler over the same depth. It seems probable, however, that the uranium depletion effect is related to $CaCO_3$ precipitation, since the geochemistry of uranium is known to be linked closely to carbonate in solutions of seawater composition[10,23,24].

Uranium concentrations in the two *in situ* sampler bottom water samples (at 1 cm and 50 cm above the sediment/water interface) are similar to the accepted uranium seawater value, with the weighted mean porewater value (3.20 ± 0.5 $\mu g\,l^{-1}$) being very slightly lower but still within the uncertainty of the seawater value. No systematic trend can be discerned in the *in situ* profile, and it therefore appears that uranium behaves

© 1984 Nature Publishing Group

Table 1 Porewater analytical data for uranium and alkalinity

Station 10552

Box core (9BX: 4,655 m: 19°27′ N 29°54′ W) Depth (cm)	Uranium ($\mu g\,l^{-1}$)	Alkalinity (mequiv. l^{-1})
0–2.5	2.14±0.09	—
5–7.5	2.04±0.09	—
10–12.5	1.99±0.08	—
15–17.5	1.72±0.08	—
20–24	1.77±0.08	—
28–32	1.50±0.07	—
32–37	1.48±0.07	—

Kastenlot core (2K: 4,683 m: 19°23′ N 29°54′ W) Depth (cm)	Uranium ($\mu g\,l^{-1}$)	Alkalinity (mequiv. l^{-1})
8–13	1.75±0.08	1.99
26–31	1.57±0.07	1.81
44–49	2.25±0.09	1.91
57–62	1.57±0.07	2.03
68–73	1.29±0.06	2.09
84–89	1.20±0.06	2.11
103–108	1.38±0.06	2.09
119–124	1.85±0.08	1.90
132–137	1.03±0.05	2.36
145–150	1.28±0.06	2.12
163–168	1.92±0.08	2.09

In situ sampler (No. 7: 4,735 m: 19°25′ N 29°53′ W) Depth (cm)	Uranium ($\mu g\,l^{-1}$)	Alkalinity (mequiv. l^{-1})
−50	3.38±0.12	2.32
−1	3.41±0.13	2.34
2	3.48±0.13	2.49
3.5	2.90±0.11	2.35
6.5	3.16±0.12	2.55
12.5	3.28±0.12	2.58
24.5	2.74±0.11	2.62
45.5	4.04±0.14	2.74

Station 10554

Box core (5BX: 5,370 m: 31°30′ N 24°26′ W) Depth (cm)	Uranium ($\mu g\,l^{-1}$)	Alkalinity (mequiv. l^{-1})
0–2	2.75±0.10	1.89
2–4	2.48±0.09	—
4–6	2.44±0.09	1.81
6–8	3.03±0.11	—
8–10	2.23±0.08	1.80
12–14.5	2.20±0.08	—
14.5–17	3.37±0.12	1.92
17–19.5	2.24±0.08	—
19.5–22	2.20±0.09	1.85
22–24.5	2.22±0.09	—
24.5–27	2.31±0.09	2.04
27–29.5	1.66±0.07	—

Kastenlot core (2K: 5,370 m: 31°30′ N 24°29′ W) Depth (cm)	Uranium ($\mu g\,l^{-1}$)	Alkalinity (mequiv. l^{-1})
21–26	14.63±0.48	1.93
35–40	10.84±0.37	2.43
53–58	10.92±0.37	2.50
76–81	11.34±0.38	2.55
95–100	8.20±0.29	—
113–118	7.60±0.28	2.59
138–143	4.45±0.18	2.48
157–162	5.63±0.22	—
184–189	4.75±0.19	2.27

In situ sampler (No. 12: 5,371 m: 31°27′ N 24°27′ W) Depth (cm)	Uranium ($\mu g\,l^{-1}$)	Alkalinity (mequiv. l^{-1})
−50	3.78±0.13	2.25
−1	2.37±0.08	2.33
2	4.96±0.17	2.63
3.5	5.36±0.18	2.51
6.5	6.43±0.24	2.79
12.5	8.76±0.31	2.92
24.5	8.69±0.31	3.03

Station 10164

Box core (Nares Abyssal Plain) (5BX: 5,613 m: 26°12′ N, 60°24′ W) Depth (cm)	Uranium ($\mu g\,l^{-1}$)
0–5	2.82±0.14
5–10	2.45±0.13
10–15	2.25±0.12
15–20	2.25±0.12
20–25	1.83±0.11
25–30	2.11±0.12

Station 10189

Box core (Bauer Deep) (4BX: 4,445 m: 9°58′ S, 102°34′ W) Depth (cm)	Uranium ($\mu g\,l^{-1}$)
0–5	2.73±0.15
5–8	2.55±0.14
8–11	2.46±0.14
11–14	2.23±0.13
14–17	2.49±0.14
17–20	2.35±0.13
20–23	2.20±0.13
23–26	2.79±0.15
26–29	2.57±0.14
29–32	3.22±0.16
35–39	2.09±0.12
42–45	2.41±0.13

conservatively in the porewaters of this slowly-accumulated carbonate sediment.

The above data relate to an oxic, carbonate sediment. At present there are no *in situ* porewater uranium profiles for carbonate-poor pelagic sediments, but squeezed box core samples from the Nares Abyssal Plain and the Bauer Deep (stations 10164 and 10189 respectively in Table 1) show uranium depletions relative to seawater similar to those observed in the cores from station 10552. There is good geochemical evidence that both of these low carbonate sediments accumulated at a rate <0.5 cm per 10^3 yr and that the porewaters remain oxic. Previously the interpretation has been that uranium removal from solution occurs within the upper 10 cm of the sediment column[5]. The finite uranium content of authigenic micronodules[25] and phillipsites[26] and the high post-depositional content of fish teeth apatite[27] illustrate that uranium uptake from solution can occur in slowly-accumulated sediments in some circumstances. Nevertheless, the demonstration here that *in situ* sampling is necessary to obtain reliable porewater uranium values means that profiles in non-calcareous sediments like those from the Nares Abyssal Plain and Bauer Deep now require redetermination on *in situ* samples before quantitative interpretation.

So far only oxic sediments have been discussed. The other station where comparative corer and *in situ* sampler data are available is more complex, because the porewaters are generally anoxic at shallow depth. The area is the abyssal plain east of Great Meteor Seamount, where pelagic sedimentation is interrupted sporadically by deposition of calcareous turbidite units of varying thickness[28]. At station 10554, the porewaters are anoxic within about 40 cm of the interface, with anoxia coincident with a 20 cm thick green turbidite centred at approximately 40 cm which has a mean organic carbon content of 1.4% and a mean uranium content of 6.1 p.p.m. Interpretation is complicated by an unexpected variability between cores taken only a few miles apart on this abyssal plain. Various measurements made on the Kastenlot core (10554 2K) yield results which differ significantly from those obtained from three other profiles at this site: the uranium data from this core are included in Fig. 3 but are not discussed further.

For the box core and *in situ* sampler (Fig. 3), the alkalinity pressure artefact is again observed. The porewater oxygen profiles are similar but not identical, both falling to <10% of air saturation at 30 cm and to zero at 40 cm (only box core data available). It is evident that the box core samples in this zone of falling oxygen tension exhibit depressed uranium porewater levels (weighted mean $2.32 \pm 0.03\ \mu g\,l^{-1}$) relative to seawater. By contrast the *in situ* sampler uranium profile is enhanced over the seawater value, and from the gradient of the uppermost four points a uranium flux from the sediments to seawater is indicated. Assuming transport of uranium as the uranyl ion, Fick's relationship[29] applies, that is, $F = -D(dC/dx)$, where F

© 1984 Nature Publishing Group

is the flux in $g\,cm^{-2}\,yr^{-1}$, D is the effective diffusion coefficient in $cm^2\,yr^{-1}$ and dC/dx is the concentration gradient in $g\,cm^{-4}$. Using for D a value of $30\,cm^2\,yr^{-1}$ which includes the effects of tortuosity and porosity[4,30], an upward flux of $1.1\times10^{-8}\,g\,cm^{-2}\,yr^{-1}$ within the porewaters is estimated. The source of this flux is believed to be the turbidite layer, with the uranium either released on oxidation of organic matter, mobilized by organic ligand complexations, or remobilized by the Eh and pH change accompanying anoxia. It is not at present possible to define the uranium speciation in natural solution/solid interactions in detail. Although considerable progress has been made in the thermodynamics of solution/mineral equilibria[10], such work has been concerned with uranium minerals and high laboratory concentrations. These contrast with the natural situation, and in particular the effects of formation of organic complexes have not been elucidated[23].

The above information suggests that the literature data on deep-sea porewater uranium contents must be regarded with caution. The systematic error observed in core samples at station 10552 demonstrates that *in situ* sampling is necessary to obtain reliable values in oxic porewaters, at least for carbonate sediments. The more complex redox situation at station 10554 is less easy to interpret, but it is evident that the core sampling artefact also appears operative in the oxic section at this station. Sampling procedures should in future be chosen to elucidate and guard against possible pressure, temperature and oxidative effects. This is the first report of a pressure artefact in the porewater sampling of deep ocean sediments for any trace element. There is at present no information on the possibility of similar artefacts for other trace elements for which carbonate speciation is important.

We thank A. B. Mackenzie for providing neutron irradiation facilities, A. S. G. Curtis and the Wellcome Trust for access to a Quantimet image analyser and NERC for studentship support (to J. Toole). The research at IOS was carried out under contract for the Department of the Environment, as part of its radioactive waste management research programme. The results will be used in the formulation of government policy, but at this stage they do not necessarily represent government policy.

Received 1 August; accepted 8 December 1983.

1. Baturin, G. N. & Kochenov, A. V. *Geokhimiya* **10**, 1529–1536 (1973).
2. Boulad, A. P. & Michard, G. *Earth planet. Sci. Lett.* **32**, 77–83 (1976).
3. Somayajulu, B. L. K. & Church, T. M. *J. geophys. Res.* **78**, 4529–4531 (1973).
4. Ku, T.-L., Knauss, K. G. & Mathieu, G. G. *Deep-Sea Res.* **24**, 1005–1017 (1977).
5. Cochran, J. K. & Krishnaswami, S. *Am. J. Sci.* **280**, 849–889 (1980).
6. Wilson, J. D., Webster, R. K., Milner, G. W. C., Barnett, G. A. & Smales, A. A. *Analyt. chim. Acta* **23**, 505–514 (1960).
7. Turekian, K. K. & Chan, L. H. in *Activation Analysis in Geochemistry and Cosmochemistry* (eds Brunfeld, A. O. & Steinnes, E.) 311–320 (Universitetsforlaget, Oslo, 1971).
8. Burton, J. D. in *Chemical Oceanography* Vol. 3, 2nd edn (eds Riley, J. P. & Skirrow, G.) 91–191 (Academic, London, 1975).
9. Starik, E. & Kolyadin, L. B. *Geokhimiya* **3**, 204–213 (1957).
10. Langmuir, D. *Geochim. cosmochim. Acta* **42**, 547–569 (1978).
11. Cochran, J. K. in *Uranium Series Disequilibrium: Applications to Environmental Problems* (eds Ivanovich, M. & Harmon, R. S.) 384–430 (Clarendon, Oxford, 1982).
12. Sayles, F. L., Mangelsdorf, P. C., Wilson, T. R. S. & Hume, D. N. *Deep-Sea Res.* **23**, 259–264 (1976).
13. Peters, R. D., Timmins, N. T., Calvert, S. E. & Morris, R. J. *Inst. Oceanogr. Sci. Rep. 106* (1980).
14. Crawford, R. W. & Baxter, M. S. *Proc. IAEA Conf. Meth. Low-level Counting Spectrometry* 263–276 (IAEA, Vienna, 1981).
15. Fleischer, R. L. & Delaney, A. C. *Analyt. Chem.* **48**, 642–645 (1976).
16. Fleischer, R. L. & Lovett, D. B. *Geochim. cosmochim. Acta* **32**, 1126–1129 (1968).
17. Sørensen, J. & Wilson, T. R. S. *Limnol. Oceanogr.* (in the press).
18. Emerson, S. *et al. Earth planet. Sci. Lett.* **49**, 57–80 (1980).
19. Sayles, F. *Geochim. cosmochim. Acta* **45**, 1061–1086 (1981).
20. Murray, J. W., Emerson, S. & Jahnke, R. *Geochim. cosmochim. Acta* **44**, 963–972 (1980).
21. Emerson, S., Grundmanis, V. & Graham, D. *Earth planet. Sci. Lett.* **61**, 220–232 (1982).
22. Jahnke, R., Heggie, D., Emerson, S. & Grundmanis, V. *Earth planet. Sci. Lett.* **61**, 233–256 (1982).
23. Giblin, A. M., Batts, B. D. & Swaine, D. J. *Geochim. cosmochim. Acta* **45**, 699–709 (1981).
24. Turner, D. R., Whitfield, M. & Dickson, A. G. *Geochim. cosmochim. Acta* **45**, 855–881 (1981).
25. Immel, R. & Osmond, J. K. *Chem. Geol.* **18**, 263–272 (1976).
26. Bernat, M., Bieri, R. H., Koide, M., Griffin, J. J. & Goldberg, E. D. *Geochim. cosmochim. Acta* **34**, 1053–1071 (1970).
27. Bernat, M. & Allegre, C. J. *Earth planet. Sci. Lett.* **21**, 310–314 (1974).
28. Weaver, P. P. E. & Kuijpers, A. *Nature* **306**, 360–363 (1983).
29. Crank, J. *The Mathematics of Diffusion* (Clarendon, Oxford, 1956).
30. Li, Y.-H. & Gregory, S. *Geochim. cosmochim. Acta* **38**, 703–714 (1978).

Colloid stability of iron oxide particles from a freshwater lake

E. Tipping & M. Ohnstad

Freshwater Biological Association, The Ferry House, Ambleside, Cumbria LA22 0LP, UK

Direct measurements of the aggregation rates of natural particles can help to explain the behaviour of material suspended in natural waters, and are a stringent test of predictions from model studies. The iron oxide which accumulates each summer in Esthwaite Water in Cumbria, UK is a natural colloid consisting of a sufficiently homogeneous population of particles to allow such direct measurements to be made. Here we show that, at initial particle concentrations (n_0) characteristic of those in the lake (10^{10}–$10^{11}\,dm^{-3}$), the colloid stability of the iron oxide is governed by the opposing forces of van der Waals' attraction and electrostatic repulsion, as predicted from model experiments on synthetic haematite particles coated with aquatic humic substances. However, at higher particle concentrations and with Ca^{2+} concentrations greater than 0.01 mol dm^{-3} it is found that the aggregation rate constant depends on n_0, possibly indicating the participation of bridging flocculation in the aggregation process under these conditions.

The aggregation of colloidally sized particles in natural waters has a major influence on material transport and the properties of cohesive sediments[1,2]. It is generally assumed that aggregation is related to the surface properties—especially the electrical potential—of the particles[3,4], but hitherto the nature of the relationship has only been investigated with model systems[5–8] or by filtration experiments with uncharacterized heterogeneous natural samples[9,10]. In the work reported here we studied the colloid stability of naturally occurring iron oxide particles by measuring aggregation rate constants (k) and electrophoretic mobilities (u) at different concentrations of calcium chloride.

Iron oxide accumulates at depth in Esthwaite Water during the early part of the lake's annual period of deep-water anoxia, to reach concentrations of Fe of about 5×10^{-5} mol dm^{-3} (refs 11, 12). The oxide is in the form of amorphous spherical, ellipsoidal or rod-shaped particles, dimensions of individual particles being in the range 0.05–0.5 μm (ref. 13). (In this work the mean diameter was 0.3 μm, with a coefficient of variation of 45%.) The particulate material is 30–40% by weight Fe, and approximately 30% by weight organic matter, of which about one-third consists of humic substances[13]. From studies with synthetic iron oxides it seems that the surface properties of the particles, as reflected by their electrophoretic mobilities, are dominated by the humic substances[14,15]. Esthwaite Water humic substances have molecular weights ~5,000 and have about 0.01 mol g^{-1} of ionizable functional groups[16].

Aggregation was followed by particle counting as described previously[8], the method involving visualization of the particles by dark-field microscopy. It was found that after dissolution of the iron oxide particles with oxalate, the number of visible particles left was approximately 5% of the original number, indicating that the results obtained in the aggregation experiments refer almost exclusively to iron oxide particles. In order to carry out an experiment, the chosen suspension was sonicated, the desired amount of $CaCl_2$ added, and aliquots taken for particle enumeration at appropriate times. Experiments were done at room temperature (20–25 °C) and at pH values close to those found for Esthwaite Water (6.6–6.9). Judging from scanning electron microscopic observation of material collected on 0.01-μm Nuclepore filters, the sonication procedure did not break up all aggregates present. However, there were very few aggregates larger than 2 μm after sonication. The suspensions used in the experiments were either lakewater samples, untreated except for sonication, or particulate matter concentrated by centrifugation or gravitational settling and diluted as

© 1984 Nature Publishing Group

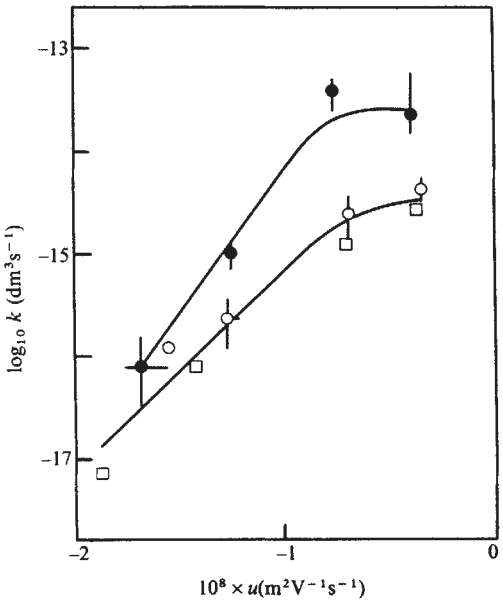


Fig. 1 Dependence of aggregation rate constant (k) on electrophoretic mobility (u) for Esthwaite Water iron oxide particles at initial concentrations (n_0) of $3–8\times10^{10}$ dm^{-3} (○) and $2–4\times10^{11}$ dm^{-3} (●), and for haematite particles ($n_0=6–9\times10^{11}$ dm^{-3}) equilibrated with surface Esthwaite lakewater (□). Each plot consists of four points: (1) (most negative u) lakewater, [Ca+Mg]≃ 3×10^{-4} mol dm^{-3}; (2) lakewater+10^{-3} mol dm^{-3} $CaCl_2$; (3) lakewater+0.01 mol dm^{-3} $CaCl_2$; (4) (least negative u) lakewater+0.1 mol dm^{-3} $CaCl_2$. For the natural oxide each point is the mean of values obtained for 2, 3 or 4 different samples; bars indicate ranges of values. In experiments at low natural concentrations (n_0) there was no systematic difference between k values obtained with unconcentrated samples and values for samples which had been concentrated and then re-diluted. Haematite gives k values independent of n_0 (see Fig. 2). The lines are drawn for guidance.

required with filtered lakewater. As well as natural oxides we also used, for comparison, ellipsoidal haematite (α-Fe_2O_3) particles prepared by a method[17] which yielded a narrow size range (mean diameter 0.4 μm, coefficient of variation 10%). The particles were equilibrated by repeated centrifugation and resuspension, with filtered (0.1 μm) surface Esthwaite lakewater sampled in April 1983 (HS concentration ~2 mg dm^{-3}); this treatment gave the particles an adsorbed layer of humic substances[14,15]. Electrophoretic mobilities were measured with a Rank Brothers Mark II apparatus at 25 °C using a flat cell.

Our experiments were carried out on unstirred suspensions, therefore aggregation was brought about by diffusion and the appropriate equation for the process is second order:

$$1/n = 1/n_0 + kt$$

where n and n_0 are the particle concentrations at time t and at zero time, respectively. The aggregation rate constant k could alternatively be expressed as k_{max} α where k_{max} is the rate constant for the aggregation of completely destabilized particles[18] and α is the collision frequency factor[19]. Plots of $1/n$ against t were reasonably linear ($r \geq 0.9$) for at least two halvings of n_0. Because a manual counting method was used, the total number of counts for an individual aliquot was quite small (100–400); this meant that the standard errors of the slopes of the plots were quite large: 10–20% of k in most cases, but up to 40% (±0.2 $\log_{10}$ units) at $k=10^{-16}$–10^{-17} dm^3 s^{-1}. However, as k varies over several orders of magnitude, depending on $CaCl_2$ concentration ([$CaCl_2$]), errors of this kind are acceptable for our present purposes.

Figure 1 shows the dependence of k on u for the Esthwaite Water iron oxide particles and for haematite particles equilibrated with surface Esthwaite lakewater. The negative values of u are due to those functional groups of the adsorbed humic substances that are not involved in the actual oxide–humic

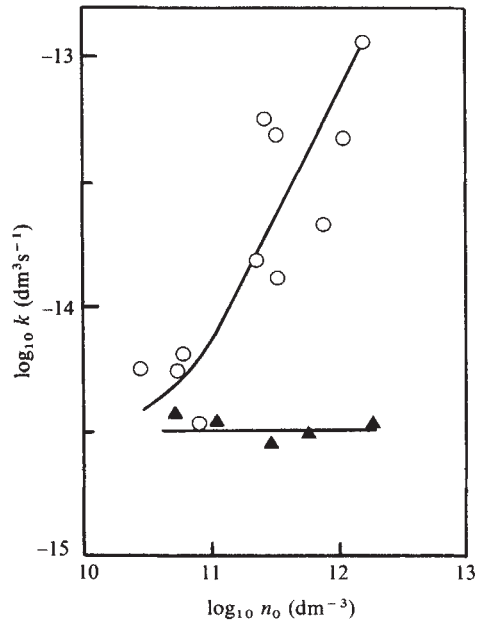


Fig. 2 Dependence of aggregation rate constant (k) on the initial particle concentration (n_0) for samples of Esthwaite Water iron oxide (○) and haematite (▲), at $CaCl_2$ = 0.1 mol dm^{-3}.

interaction[15]. The magnitudes of u for the natural and model particles differ somewhat, especially at low [$CaCl_2$]. This may reflect slight differences in major ion and humic concentrations in the suspensions, the haematite particles being suspended in filtered surface Esthwaite lakewater sampled in April, while the natural oxide particles were suspended in water from 1 m above the sediment sampled in June or July. Also, at least for Mn oxides, the magnitude of the mobility of humic-coated particles depends to a small extent on the nature of the underlying surface[20]. The electrophoretic mobilities become less negative with increasing [$CaCl_2$] due to the specific interactions of Ca^{2+} with the humic functional groups in the electrokinetic shear plane[15].

The dependence of k on u for the Esthwaite Water oxide particles at their natural lakewater concentrations ($3–8\times10^{10}$ dm^{-3} after sonication) is similar to that found for the humic-coated haematite particles (Fig. 1). It seems that for both types of particle the aggregation rate depends on the degree of interaction of Ca^{2+} with the adsorbed humics, so that it is inversely related to the interparticle electrostatic repulsion. Furthermore, the plot in Fig. 1 indicates a maximum value of $k\sim4\times10^{-15}$ dm^3 s^{-1}, which is close to the theoretical maximum value for diffusion-controlled aggregation (corrected for hydrodynamic effects[21]) of completely destabilized spheres, that is 3.2×10^{-15} dm^3 s^{-1} at 25 °C (ref. 22). It thus appears that the colloid stability of the particles in these conditions is governed simply by two opposing forces, van der Waals' attraction and electrostatic repulsion (compare with ref. 23). At high calcium concentration, only the attractive force operates.

Before discussing aggregation at higher n_0 it is appropriate to point out that the results for the Esthwaite Water system refer to relatively low humic adsorption densities. In previous model studies[8] with humic-coated haematite particles it was found that high adsorption densities could give rise to steric stabilization. Thus, in 0.1 mol dm^{-3} $CaCl_2$, k was only 5×10^{-16} dm^3 s^{-1} in the presence of 20 mg HS dm^{-3} compared with 3×10^{-15} dm^3 s^{-1} at 1 mg HS dm^{-3}. Therefore, in Esthwaite Water the humics influence colloid stability only via electrostatic effects, but in more humic-rich waters an additional steric influence is to be expected.

At $n_0=2–4\times10^{11}$ dm^{-3} (that is, after concentration) the dependence of k on u for the Esthwaite Water particles differs from that at lower n_0 and for haematite (Fig. 1), with aggregation at [$CaCl_2$] ≥ 0.01 mol dm^{-3} being accelerated to become substantially faster than the theoretical maximum referred to above.

© 1984 Nature Publishing Group

Figure 2 shows that at $[CaCl_2]$ of 0.1 mol dm^{-3}, k increases by an order of magnitude as n_0 increases from 10^{11} to 10^{12} dm^{-3}. In contrast, the haematite particles suspended in filtered surface Esthwaite lakewater give essentially constant values of k with n_0.

Thus at $n_0 \geq 10^{11}$ dm^{-3} it seems that a second factor, in addition to charge destabilization, is involved in causing the natural iron oxide particles to aggregate. This second factor does not apply to the haematite particles and so cannot be related to the relatively minor increase in k (~three-fold between $n_0 = 10^{10}$ and $n_0 = 2.5 \times 10^{12}$ dm^{-3}) found by Hatton *et al.*[22] in their study of destabilized polystyrene latex spheres, the enhanced k possibly being due to increasing multiparticle interactions with increasing n_0. Furthermore, polydispersity of the natural particles can be ruled out as an explanation of the high k values (compare ref. 24), first because very high polydispersity is required for order-of-magnitude changes in k, and second because the effect should be independent of n_0.

The most plausible explanation of the results at high n_0 seems to us to be some kind of bridging flocculation (see, for example, ref. 25). Enhancement of aggregation rates above the maximum rate for salt-induced coagulation has been observed for dispersions of silver iodide with polyacrylic acid[26], cellulose with alginates[27] and polystyrene latex with proteins[28], in each case the enhancement being attributed to bridging interactions by the polymers. With regard to the increase of k with n_0, Walles[29] has published a theory, based on an enlarged effective collision diameter for particles with attached polymers, which predicts such an effect for bridging flocculation. Our observation that the enhancement of k is most marked at higher $CaCl_2$ concentrations could be explained by the bridging being more effective as the electrical double layers of the interacting particles became more suppressed[25]. Alternatively, if the bridging agents are not integral parts of the oxide particles, Ca^{2+} may induce their interaction (compare with refs 27, 30).

There are several candidates for the bridging agent(s); algal or bacterial exudates[27,31,32], very high molecular weight humic substances, and organic fibrils[33]. The total lack of dependence of k on n_0 for the haematite particles (Fig. 2) means either that the putative bridging agents are absent from the filtered surface water, or that they are present but do not interact with the haematite to bring about accelerated aggregation rates.

The possible bridging flocculation considered above will not apply to freshwater aggregation processes because of the requirement for high concentrations of Ca^{2+}, but it could be important for aggregation in estuaries, notably at the turbidity maximum. It remains to be seen whether the dependence of k on n_0, and possible bridging flocculation, is a widespread phenomenon for natural particles, or whether the more straightforward behaviour we observe at low n_0 is more common. In either event, the results presented here, together with the previous finding[8] that high adsorption densities of humic substances can enhance colloid stability by steric stabilization (see above), suggest that detailed understanding of aggregation in natural waters requires more direct studies of naturally occurring particles of various types.

We thank J. Hilton, P. S. Liss, L. M. Mayer and D. J. A. Williams for helpful comments and J. C. Rhodes for typing the manuscript. This work was supported by the NERC.

Received 6 October; accepted 7 December 1983.

1. Lerman, A. *Geochemical Processes* (Wiley, New York, 1979).
2. Yariv, S. & Cross, H. *Geochemistry of Colloid Systems* (Springer, Berlin, 1979).
3. Stumm, W. & Morgan, J. J. *Aquatic Chemistry* 2nd edn (Wiley, New York, 1981).
4. Hunter, K. A. & Liss, P. S. *Nature* **282**, 823–825 (1979).
5. Edzwald, J. K., Upchurch, J. B. & O'Melia, C. R. *Envir. Sci. Technol.* **8**, 58–63 (1974).
6. Hahn, H. H. & Eppler, B. in *Colloid and Interface Science* Vol. 4 (ed. Kerker, M.) 125–137 (Academic, New York, 1976).
7. Hunt, J. R. *Envir. Sci. Technol.* **16**, 303–309 (1982).
8. Tipping, E. & Higgins, D. C. *Colloid Surf.* **5**, 85–92 (1982).
9. Sholkovitz, E. R. *Geochim. cosmochim. Acta* **40**, 831–845 (1976).
10. Mayer, L. M. *Geochim. cosmochim. Acta* **46**, 2527–2535 (1982).
11. Mortimer, C. H. *J. Ecol.* **29**, 280–329 (1941); **30**, 147–201 (1942).
12. Davison, W., Heaney, S. I., Talling, J. F. & Rigg, E. *Schweitz. Z. Hydrol.* **42**, 196–224 (1981).
13. Tipping, E., Cook, D. & Woof, C. *Geochim. cosmochim. Acta* **45**, 1411–1419 (1981).
14. Tipping, E. *Geochim. cosmochim. Acta* **45**, 191–199 (1981).
15. Tipping, E. & Cooke, D. *Geochim. cosmochim. Acta* **46**, 75–80 (1982).
16. Tipping, E. *Chem. Geol.* **33**, 81–89 (1981).
17. Matijevic, E. & Scheiner, P. *J. Colloid Interface Sci.* **63**, 509–524 (1978).
18. Von Smoluchowski, M. *Z. phys. Chem.* **92**, 129–168 (1918).
19. O'Melia, C. R. *Envir. Sci. Technol.* **14**, 1052–1060 (1980).
20. Tipping, E. & Heaton, M. J. *Geochim. cosmochim. Acta* **47**, 1393–1397 (1983).
21. Spielman, L. A. *J. Colloid Interface Sci.* **33**, 562–571 (1970).
22. Hatton, W., McFadyen, P. & Smith, A. L. *JCS Faraday* **70**, 655–660 (1974).
23. Overbeek, J. T. G. in *Colloid Science* Vol. 1 (ed. Kruyt, H. R.) 302–341 (Elsevier, Amsterdam, 1952).
24. Müller, H. *Kolloidchem. Beih.* **26**, 257–311 (1928).
25. Gregory, J. in *The Scientific Basis of Flocculation* (ed. Ives, K. J.) 101–130 (Sijthoff & Noordhoff, Alphen aan den Rijn, 1978).
26. Williams, D. J. A. & Ottewill, R. H. *Kolloid-Z. Z. Polym.* **243**, 141–147 (1971).
27. Lindström, T. & Söremark, C. *J. Colloid Interface Sci.* **55**, 69–72 (1976).
28. Van der Scheer, A., Tanke, M. A. & Smolders, C. A. *Discuss. Faraday Soc.* **65**, 264–281 (1978).
29. Walles, W. E. *J. Colloid Interface Sci.* **27**, 797–803 (1968).
30. Sommerauer, A., Sussman, D. L. & Stumm, W. *Kolloid-Z. Z. Polym.* **225**, 147–154 (1968).
31. Busch, P. L. & Stumm, W. *Envir. Sci. Technol.* **2**, 49–53 (1968).
32. Eaton, A. *Est. Coast. mar. Sci.* **9**, 41–49 (1979).
33. Massalaski, A. & Leppard, G. G. *J. Fish. Res. Bd Can.* **36**, 906–921 (1979).

^{230}Th–^{238}U disequilibrium in island arcs: evidence from the Aleutians and the Marianas

S. Newman*, J. D. Macdougall & R. C. Finkel*

Geological Research Division, A-020, Scripps Institution of Oceanography, La Jolla, California 92093, USA

Complex models for magma generation in island arcs have recently been proposed which require varying contributions from three possible sources: (1) the mantle; (2) subducted, altered oceanic crust; and (3) sediments (see, for example, refs 1–6). These models involve partial melting of metasomatized mantle. The metasomatizing liquids derive from the subducted slab and are large-ion-lithophile (LIL) element-bearing fluids or are themselves partial melts. Further modification of the chemical composition of the lavas may occur by processes such as shallow-level fractional crystallization. Because the Th/U ratio is quite different for each of the three possible magma sources mentioned above, and as Th and U may behave differently during fractionation processes, particularly if volatile transport is involved, studies of ^{230}Th–^{238}U disequilibrium[7–11] have great potential in investigations of arc lavas. Our data, given here, imply that the Th/U ratio in the sources of Mariana and Aleutian Island Arc lavas is similar to ocean ridge basalt sources, although there is a hint of a somewhat lower Th/U component in the Aleutian source. The data also confirm and extend the observations by Gill[5] and Allègre and Condomines[10] that oceanic arc lavas are distinguished from magmas erupted in virtually all other volcanic environments by activity ratios ($^{230}Th/^{238}U$) < 1, implying relative enrichment of uranium.

The ^{230}Th–^{238}U disequilibrium system is a useful tool for studying sources and processes involved in recent volcanism because it can provide a direct measure of an important geochemical characteristic of the source, the Th/U ratio[7–10]. Unlike other isotopic tracers such as $^{87}Sr/^{86}Sr$, $^{143}Nd/^{144}Nd$, or $^{206}Pb/^{204}Pb$, which reflect the relevant parent–daughter ratios integrated over the lifetime of the mantle, the activity ratio ($^{230}Th/^{232}Th$) reflects the current U/Th in the source (activities are indicated by parentheses throughout this paper; activity = λN, where λ is the appropriate decay constant). This is so because ^{230}Th is radioactive with a half life of 75,200 years, much less than its ultimate parent ^{238}U. Thus radioactive equilibrium is attained rapidly (~300,000 yr), a state in which (^{230}Th) = (^{238}U), that is, ($^{230}Th/^{232}Th$) = ($^{238}U/^{232}Th$). Therefore if the source is in equilibrium and the transfer time between source and surface is short, the initial ($^{230}Th/^{232}Th$) of a lava will reflect U/Th in the source. In contrast, since petrogenetic processes may lead to fractionation between Th and U, ($^{230}Th/^{238}U$) in the lava may not reflect the source, that is, ($^{230}Th/^{238}U$) ≠ 1.0. When this occurs, data points fall away from the equiline on plots such as Fig. 3.

* Present addresses: Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA (S.N.); Ice Core Laboratory, State University of New York, Buffalo, 4240 Ridge Lea Road, Amherst, New York, 14226, USA (R.C.F.).

© 1984 Nature Publishing Group

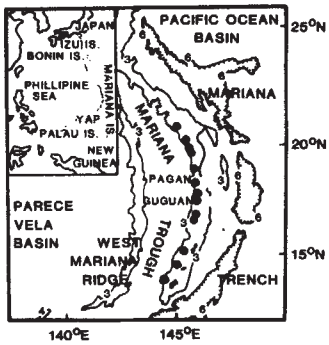


Fig. 1 Location map for the Mariana samples. Two samples from Pagan and one from Guguan were analysed. Contours are in km (adapted from ref. 28).

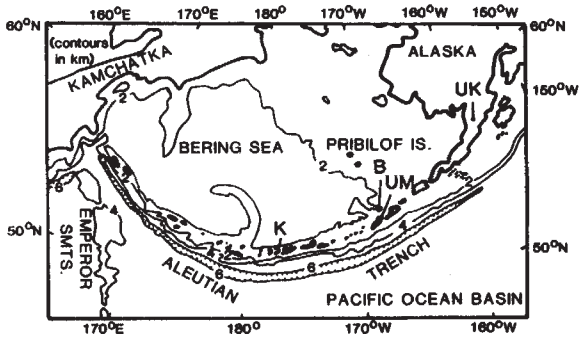


Fig. 2 Location map for the Aleutian samples. K, Kanaga; UM, Umnak; B, Bogoslof; UK, Ukinrek Maar. Contours are in km (adapted from ref. 29).

Our Th–U data for the Mariana and Aleutian samples are reported in Table 1; details of the analytical methods are presented elsewhere[11,12]. All samples are of negligible age with respect to the half life of ^{230}Th. The Mariana samples are from Pagan and Guguan Islands (Fig. 1); the Aleutian lavas are from Kanaga, Umnak and Bogoslof Islands and Ukinrek Maar on the Alaskan peninsula (Fig. 2). All samples analysed are basalts, except for the 1796 lava from Bogoslof, which is an andesite.

For each arc, (^{230}Th/^{232}Th) is approximately constant (Fig. 3). This suggests that each has a source with approximately constant Th/U. However, more analyses will be required to confirm this suggestion for the Marianas, as only three samples from two islands were analysed. Dixon and Batiza[13] have suggested that the low rate of eruption, together with evidence of significant shallow fractional crystallization in the Marianas (see also ref. 14), imply long residence in sub-volcanic reservoirs. However, this residence time appears to be insignificant with respect to ^{230}Th decay, since it would be coincidental for lavas from two islands, Pagan and Guguan, to have experienced varying residence times resulting in the same (^{230}Th/^{232}Th)—within the 2σ uncertainty—at eruption. Because (^{230}Th/^{238}U) < 1, residence in a magma chamber would increase (^{230}Th/^{232}Th) relative to the source. The Mariana thorium isotopic ratios

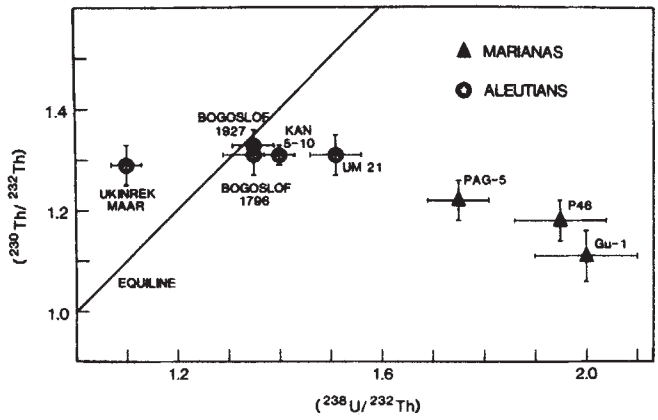


Fig. 3 (^{230}Th/^{232}Th) versus (^{238}U/^{232}Th) for the Mariana and Aleutian lavas. Error bars indicate 1σ counting statistics uncertainties. The equiline represents the locus of points in radioactive equilibrium, for which (^{230}Th) = (^{238}U).

(average (^{230}Th/^{232}Th) = 1.17 ± 0.06) correspond to Th/U of ~2.6, similar to or slightly higher in Th/U relative to mid-ocean ridge basalt (MORB) source[8,15].

The thorium isotopic compositions of the five Aleutian lavas from four volcanoes are essentially constant at 1.31 ± 0.02 (Fig. 3) despite the transition from continental to oceanic crust at approximately the position of Bogoslof (Fig. 2) and the change in differentiation trend from tholeiitic (Umnak) to calc-alkaline (Kanaga) to alkalic (Bogoslof and Ukinrek Maar). It is unlikely that lavas from four widely separated volcanoes would have the same isotopic compositions unless the (^{230}Th/^{232}Th) ratios reflect sources with approximately constant Th/U. The inferred Th/U of 2.35 is slightly lower than that in the average MORB source, and while a few oceanic basalts with high (^{230}Th/^{232}Th) approach the Aleutian average, it nevertheless appears to be distinct.

Based on the thorium isotopic data, the source regions for both the Mariana and Aleutian lavas appear to be homogeneous with respect to Th/U. However, this ratio is different in the two sources and the difference may indicate disparate components. Indeed, other geochemical parameters suggest that lavas from these two arcs are derived from complex sources consisting of different proportions of depleted MORB-type mantle, altered oceanic crust, and sediment[2,13,14,16–20].

For the Aleutians, Kay[2] and McCulloch and Perfit[16] concluded that the continental crust underlying the eastern part of the arc has contributed little, if at all, to the lavas of the volcanoes on the Alaskan peninsula. Petrogenetic studies by Kienle *et al.*[21] on the 1977 eruptives from Ukinrek Maar, together with the observation from this study that a sample from the same flow has the same (^{230}Th/^{232}Th) as those from the intra-oceanic volcanoes, support this conclusion. The high and constant

Table 1 Th–U data for lavas from the Mariana and Aleutian Island Arcs

Sample	Th (p.p.m.)	U (p.p.m.)	Th/U	^{230}Th (d.p.m. g^{-1})	(^{234}U/^{238}U)	(^{230}Th/^{238}U)	(^{238}U/^{232}Th)	(^{230}Th/^{232}Th)	Age/date
Marianas									
Gu-1 Guguan	0.23 ± 0.01	0.146 ± 0.004	1.55 ± 0.08	0.061 ± 0.003	0.97 ± 0.03	0.56 ± 0.03	2.00 ± 0.10	1.11 ± 0.05	300 yr
PAG-5 Pagan	0.48 ± 0.01	0.272 ± 0.006	1.76 ± 0.06	0.140 ± 0.004	0.99 ± 0.03	0.70 ± 0.02	1.75 ± 0.06	1.22 ± 0.04	Spring, 1925
P46 Pagan	0.56 ± 0.02	0.355 ± 0.007	1.59 ± 0.07	0.159 ± 0.006	0.98 ± 0.02	0.60 ± 0.03	1.95 ± 0.09	1.18 ± 0.04	5/15/81
Aleutians									
KAN 5–10 Kanaga	3.73 ± 0.07	1.69 ± 0.03	2.21 ± 0.05	1.17 ± 0.02	0.99 ± 0.02	0.94 ± 0.02	1.40 ± 0.03	1.31 ± 0.02	~1900
UM 21 Umnak	1.36 ± 0.03	0.66 ± 0.01	2.06 ± 0.06	0.426 ± 0.009	1.02 ± 0.02	0.87 ± 0.02	1.51 ± 0.05	1.31 ± 0.04	~1948
B1796 Bogoslof	5.5 ± 0.2	2.43 ± 0.04	2.3 ± 0.1	1.75 ± 0.06	1.00 ± 0.02	0.97 ± 0.04	1.35 ± 0.06	1.31 ± 0.04	1796
B1927 Bogoslof	3.34 ± 0.09	1.46 ± 0.02	2.29 ± 0.07	1.06 ± 0.03	1.02 ± 0.02	0.98 ± 0.03	1.35 ± 0.04	1.33 ± 0.03	1927
Ukinrek Maar	1.58 ± 0.04	0.56 ± 0.01	2.82 ± 0.09	0.49 ± 0.01	1.01 ± 0.02	1.18 ± 0.03	1.10 ± 0.03	1.29 ± 0.04	1977

Errors quoted are 1σ counting statistics uncertainties.

© 1984 Nature Publishing Group

thorium isotopic ratio of the Aleutian lavas (1.31) compared with that of MORB (average 1.22, total range 1.15–1.30; refs 8, 15), requires a component with low Th/U. Continental or arc-derived sediments tend to have high Th/U, unless seawater alteration has resulted in uranium addition. On the other hand, carbonate sediments and old, altered MORB tend to have low Th/U, the latter because of uranium uptake from seawater. Altered oceanic crust is certainly being subducted in the Aleutian Trench, and although carbonate-rich sediments are not common in surface sediments at high latitudes, especially in the Pacific, they are present in the sediment column. Indeed, Sun[22] has reported Th/U ratios of 0.77 and 0.99 for sediments cored seaward of the Aleutian Trench. Therefore incorporation of U and Th derived from subducted crust, either in fluids which metasomatized the overlying mantle wedge or directly as partial melts, could explain the observed high ($^{230}Th/^{232}Th$) relative to MORB. A sediment contribution of 2–3% from the downgoing slab is required by the data of Brown *et al.*[23], who reported significant concentrations of ^{10}Be in Aleutian lavas. Nevertheless, it appears fortuitous that the proportions of U and Th from the subducted slab, incorporated into lavas along the length of the arc, should result in such a constant ($^{230}Th/^{232}Th$).

In contrast to data for Aleutian volcanic rocks, the Sr, Pb, Nd and He isotopic compositions of Mariana arc lavas generally rule out significant involvement of subducted sediments or altered MORB but suggest instead that the mantle source is slightly enriched in LIL elements relative to depleted, N-type MORB source[13,14,17–19]. Pb isotopic ratios[17] appear to place an upper limit of ~1% on the subducted sediment contribution. However, DePaolo and Wasserburg[20] found elevated $^{87}Sr/^{86}Sr$ ratios suggesting the presence of a seawater Sr component. The ($^{230}Th/^{232}Th$) data imply no significant incorporation of subducted sediments or altered MORB. The Mariana arc source resembles that of MORB with respect to Th/U ratio.

A very significant feature of the disequilibrium data reported here, which distinguishes the island arc environment from others studied by this technique[7–12,15], is that most samples analysed have ($^{230}Th/^{238}U$) ratios which are considerably less than 1.00, and thus fall to the right of the equiline (Fig. 3). This feature must result from relative U enrichment during petrogenesis. The tendency of subduction-related volcanism to display depletion of ^{230}Th relative to ^{238}U has been noted previously by Gill[5] and Allègre and Condomines[10]. However, our Aleutian data actually exhibit a considerable range in ($^{230}Th/^{238}U$) for constant ($^{230}Th/^{232}Th$). Lava from Ukinrek Maar exhibits ($^{230}Th/^{238}U$) > 1 while Bogoslof is characterized by equilibrium between ^{230}Th and ^{238}U. The thorium isotopic ratio of the sample from Kanaga is within 2σ, but beyond 1σ, statistical uncertainty of the equiline. The observation that the Ukinrek Maar basalt has ($^{230}Th/^{238}U$) > 1 may be related to this lava's somewhat anomalous composition (alkali olivine basalt transitional to high-alumina basalt) relative to the dominant Aleutian lava types (tholeiitic basalt to andesite). Kienle *et al.*[21] suggest that this is a primitive lava generated by a small degree of partial melting of garnet peridotite at ≥80 km depth in the mantle wedge above the subducted slab. Thus Th/U fractionation during magma generation might be expected to be different from that of other Aleutian volcanoes. However, ($^{230}Th/^{232}Th$) implies a similar Th/U in the source. Cascades lavas (ref. 11 and S. Newman, J. D. Macdougall and R. C. Finkel, in preparation) also display ($^{230}Th/^{238}U$) ≥ 1. Based on the Aleutian and Cascades data it seems that continent-based subduction-related volcanism is characterized by excess ^{230}Th or ^{230}Th in equilibrium with ^{238}U, whereas island arc volcanism is characterized by a deficiency of ^{230}Th relative to ^{238}U. Some lavas from the continental arc in Costa Rica are deficient in ^{230}Th (ref. 7), an apparent inconsistency. However, the crust underlying the Costa Rican volcanoes is actually much more similar to oceanic than continental crust[24].

Gill[5] and Allègre and Condomines[10] have proposed that the relative U enrichment in island arc environments occurs by metasomatism of the source region by fluids of either mantle or continental origin. This process must occur relatively closely in time (probably ≪ 250,000 yr) to eruption since disequilibrium is observed. Perhaps both U and Th are transported by the slab-derived fluids and Th is retained in the metasomatized mantle by accessory phases. Saunders *et al.*[25] have suggested that depletion of high field strength elements such as Zr, Ti, and Hf in island arc lavas compared to MORB could be due to "high P_{H_2O} and f_{O_2} [oxygen fugacity] conditions in the subduction zone environment", stabilizing "minor mineral phases such as ilmenite, sphene (containing Ti, Nb and Ta), zircon (containing Zr and Hf) and apatite (containing P) within the eclogite slab". Most of these minerals are known to contain large amounts of uranium and thorium. However, the limited data available suggest that these phases all have low to moderate Th/U ratios, ~0.4–3 (ref. 26). It has been suggested that amphibole may also be a residual phase and this mineral tends to have higher Th/U ratios[11,15,27]. However, high P_{H_2O} and f_{O_2} conditions in island arc sources have not been verified.

If fluid transport of U, with or without Th, explains the enrichment of U in island arc lavas, then some other process must be operative in continental arcs. Continental lavas often assimilate continental crust, during which Th is apparently more incompatible than U (ref. 11). Clearly, more data are needed on the behaviour of elements in metasomatic fluids, and on Th and U distribution coefficients for many different minerals in varying conditions, before these various possibilities can be evaluated quantitatively.

The three major conclusions which can be drawn from the ^{230}Th–^{238}U disequilibrium data for island arcs presented here are: (1) Lavas of negligible age relative to the half life of ^{230}Th from the Aleutians and the Marianas exhibit constant ($^{230}Th/^{232}Th$) ratios, implying relatively constant, but different, Th/U in the source(s) of each arc. (2) The Aleutian source probably includes a larger contribution from subducted oceanic crust than does the Mariana source. (3) Island arcs are the only environment studied so far from which lavas consistently exhibit a deficiency of ^{230}Th relative to ^{238}U, probably due to very recent preferential enrichment of the magma in U by fluids derived from the subducted slab.

The authors thank the following for providing the samples studied in this project: Drs N. G. Banks, T. H. Dixon, R. Kay and R. J. Stern. This research was supported by NSF grants EAR8000484 and EAR8026448. SIO-IG²L contribution 21.

Received 2 August; accepted 7 December 1983.

1. Ringwood, A. E. in *Island Arcs, Deep Sea Trenches, and Back-Arc Basins* (eds Talwani, M. & Pitman, W. C. III) 311–324 (American Geophysical Union, Washington, D.C., 1977).
2. Kay, R. W. *J. Geol.* **88**, 497–522 (1980).
3. Perfit, M. R., Gust, D. A., Bence, A. E., Arculus, R. J. & Taylor, S. R. *Chem. Geol.* **30**, 227–256 (1980).
4. Nicholls, I. A., Whitford, D. J., Harris, K. L. & Taylor, S. R. *Chem. Geol.* **30**, 177–199 (1980).
5. Gill, J. B. *Orogenic Andesites and Plate Tectonics* (Springer, Berlin, 1981).
6. Mysen, B. O. in *Andesites* (ed. Thrope, R. S.) 489–522 (Wiley, New York, 1982).
7. Allègre, C. J. & Condomines, M. *Earth planet. Sci. Lett.* **28**, 395–406 (1976).
8. Condomines, M., Morand, P. & Allègre, C. J. *Earth planet. Sci. Lett.* **55**, 247–256 (1981).
9. Condomines, M., Morand, P., Allègre, C. J. & Sigvaldason, G. *Earth planet. Sci. Lett.* **55**, 393–406 (1981).
10. Allègre, C. J. & Condomines, M. *Nature* **299**, 21–24 (1982).
11. Newman, S. thesis, Univ. of California, San Diego (1983).
12. Newman, S., Macdougall, J. D. & Finkel, R. C. *Geochim. cosmochim. Acta* (in the press).
13. Dixon, T. H. & Batiza, R. *Contr. miner. Petrol.* **70**, 167–181 (1979).
14. Stern, R. J. *Contr. miner. Petrol.* **68**, 207–219 (1979).
15. Newman, S., Finkel, R. C. & Macdougall, J. D. *Earth planet. Sci. Lett.* **65**, 17–33 (1983).
16. McCulloch, M. T. & Perfit, M. R. *Earth planet. Sci. Lett.* **56**, 167–179 (1981).
17. Meijer, A. *Bull. geol. Soc. Am.* **87**, 1358–1369 (1976).
18. Craig, H., Lupton, J. E. & Horibe, Y. *Adv. Earth planet. Sci.* **3**, 43–56 (1978).
19. Stern, R. J. & Bibee, L. D. *Carnegie Inst. Wash. Yb.* **79**, 465–472 (1980).
20. DePaolo, D. J. & Wasserburg, G. J. *Geophys. Res. Lett.* **4**, 465–468 (1977).
21. Kienle, J., Kyle, P. R., Self, S., Motyka, R. J. & Lorentz, V. *J. Volcanol. Geotherm. Res.* **7**, 11–37 (1980).
22. Sun, S.-S. *Phil. Trans. R. Soc.* A**297**, 409–445 (1980).
23. Brown, L., Klein, J., Middleton, R., Sacks, I. S. & Tera, F. Nature **299**, 718–720 (1982).
24. DeBoer, J. *Tectonophysics* **56**, 221–259 (1979).
25. Saunders, A. D., Tarney, J. & Weaver, S. D. *Earth planet. Sci. Lett.* **46**, 344–360 (1980).
26. Wedepohl, K. H. (ed.) *Handbook of Geochemistry* Vol. 2/5 (Springer, New York, 1978).
27. Taddeucci, A., Broecker, W. S. & Thurber, D. L. *Earth planet. Sci. Lett.* **3**, 338–342 (1967).
28. Karig, D. E. *Bull. geol. Soc. Am.* **82**, 323–344 (1971).
29. Kay, R. W., Sun, S.-S. & Lee-Hu, C.-N. *Geochim. Cosmochim. Acta* **42**, 263–273 (1978).

© 1984 Nature Publishing Group

Mastodon butchery by North American Paleo-Indians

Daniel C. Fisher

Museum of Paleontology and Department of Geological Sciences, University of Michigan, Ann Arbor, Michigan 48109, USA

It has often been argued that North American Paleo-Indians hunted both mammoths and mastodons[1–3]. However, while numerous archaeological sites involving mammoths (genus *Mammuthus*) are recognized[3,4], very few sites demonstrate direct human association with mastodons[5–8]. I report here a taphonomic analysis of several late Pleistocene mastodon (*Mammut americanum*) skeletons excavated in southern Michigan which provides compelling evidence of mastodon butchery. Butchery practices involved the production and use of tools fashioned from bones of the animal being butchered. Evidence for butchery and bone tool use includes: patterns of bone distribution and disarticulation recorded from a primary depositional context, disarticulation marks and cutmarks on bones, green bone fracturing, use wear and impact features on bone fragments, and burned bone. Moreover, determinations of the season of death of butchered mastodons[9] suggest that butchery was associated with hunting and killing, not simply scavenging of natural deaths. These findings provide new evidence of a well developed 'bone technology'[10–13] used by Paleo-Indians in eastern North America. They also add to our perception of Paleo-Indian subsistence activities and their possible role in the late Pleistocene extinction of mastodons.

The principal specimen discussed here is a partial mastodon skeleton (University of Michigan (UM) specimen 57705) excavated from a deeply stratified peat sequence near Pleasant Lake, in Washtenaw County, Michigan. Wood fragments removed from the pulp cavities of both tusks were dated at 10,395 ± 100 yr BP, and wood located stratigraphically below the skeleton was dated at 12,845 ± 165 BP using Beta-1388 and Beta-1389 (Beta Analytic Inc.). The former date is more closely associated, suggesting an age for the skeleton near the end of Paleo-Indian occupation of southern Michigan[14]. Much of the evidence for butchery at the Pleasant Lake site has been replicated by other recently excavated specimens (UM 57856 and 58028) and by specimens in older museum collections (UM 14404, 37811 and 57022). Specimens representing natural deaths have also been excavated (UM 57648 and 61246) and studied in existing collections (UM 23498 and 59936). Geological and floral evidence, as well as radiocarbon dating, suggests that all these sites are about the same age as the Pleasant Lake site.

The pattern of bone distribution at the Pleasant Lake site consists of fully articulated skeletal units displaced from one another and surrounded by isolated bones and bone fragments (Fig. 1). Direct evidence of butchery as the cause for this pattern is provided by matching marks on pairs of bones that were articulated in life but found disarticulated at the site. The conarticular surfaces of the atlas and axis vertebrae (Fig. 2) have three such pairs of marks; each mark superimposes on its counterpart as the vertebrae are rearticulated. Particularly notable are the size and character (examined by scanning electron microscopy, SEM[15]) of the highly polished, striated pair of marks C-C′ in Fig. 2. Microscopic features indicate that these marks were made by driving a relatively smooth, wedge-shaped object, probably made of bone, between the conarticular surfaces of the intact joint. Identical pairs of marks occur at other joints, notably the left knee. I have replicated all features of these marks through experiments with fresh bone and have confirmed the effectiveness of bone wedges in producing disarticulation.

Additional evidence of butchery includes a variety of cutmarks documented by SEM analysis of microstructural features[15]. These marks, made by stone and/or bone tools, can be distinguished from gnaw marks of non-human predators and scavengers, and from possible excavator's marks, on the basis of criteria developed elsewhere[16–18].

Certain bone fragments were apparently used as expedient tools[10]. Glassy polish, interpreted as use wear, is developed along formerly sharp edges of fragments that are appropriate in size and shape for cutting or scraping soft tissues. Polish is absent from comparable surfaces of other fragments. I have replicated this polish using fresh bone, and it can be distinguished from polish of non-human origin through SEM analysis[15]. Most examples of polish occur on fragments of bones that have little cancellous, marrow-bearing tissue. Their structure and anatomy make them unlikely candidates for accidental breakage or for breakage associated with marrow removal, but they are ideal for tool production. All breakage appears to have occurred while the bones were fresh, and it is accompanied by features such as local crushing at points of impact, bulbs of percussion, removal of multiple flakes from the margins of larger fragments, and broad impact depressions surrounded by sets of concentric fractures. Previous work suggests that these patterns, particularly when developed on large proboscidean bones, are indicative of the manufacture of expedient bone tools[10–13].

Evidence of burning is present on some bones. Its distribution indicates that burning occurred while much of the bone surface was still covered by soft tissue, and its microscopic features indicate attainment of surface temperatures of 440–650 °C, higher than temperatures available in most natural fires[15,19].

Analysis of incremental lamination within tusks and molars provides important life history information and also a basis for determining season of death[20]. Natural deaths analysed thus far occurred near the end of winter or the very beginning of spring,

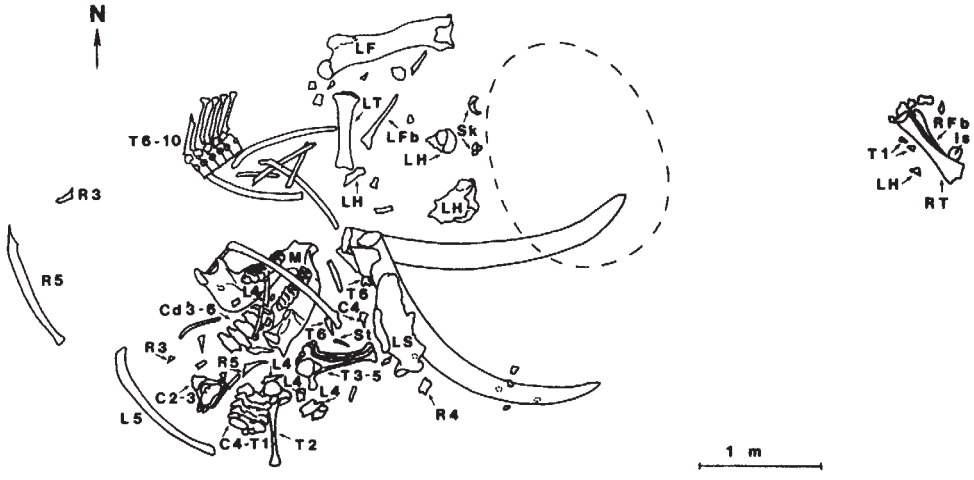


Fig. 1 Map of *in situ* bone distribution at Pleasant Lake mastodon site (UM 57705); identifications given for articulated units and representative fragments only. The dashed line demarcates the area from which the atlas vertebra and several other elements were displaced immediately before recognition of the skeleton. All other elements were undisturbed. The axis vertebra was discovered in articulation with the third cervical vertebra. C, cervical vertebra; Cd, caudal vertebra; L, left rib; LF, left femur; LFb, left fibula; LH, left humerus; LS, left scapula; ls, limestone cobble; LT, left tibia; M, mandible; R, right rib; RFb, right fibula; RT, right tibia; Sk, skull fragments; St, left stylohyoid; T, thoracid vertebra; integer suffixes indicate position in anatomical series.

© 1984 Nature Publishing Group

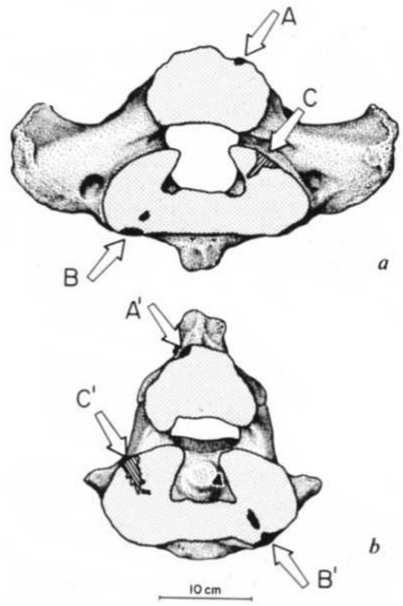


Fig. 2 Disarticulation marks on the atlas and axis vertebrae of UM 57705 (articular surfaces are indicated by even stippling). *a*, Posterior aspect of atlas, showing gouges (A and B, solid areas) and polished, striated mark (C, parallel lines indicate striae). *b*, Anterior aspect of axis, showing gouges (A′ and B′) and the polished, striated mark (C′) that are counterparts to the marks on the atlas. Superimposition occurs in the order B–A–C, as the bones are initially disarticulated and rotated slightly.

in a time of high physiological stress and heightened probability of accidental death. In contrast, all butchered animals died in mid-to-late autumn. This pattern is most plausibly explained in terms of hunting by humans[9,20].

The uniformity of patterns of bone modification at sites studied here suggests a highly standardized approach to hunting and carcass processing. This pattern shows similarities to other occurrences of modified proboscidean bone[10–13,21,22], and to certain occurrences of bison[23–25]. While lithic tools must have been used to some extent in butchery, none have been recovered from Michigan mastodon sites. Archaeological evidence of a high degree of conservation of lithic materials[26] suggests that this may be due largely to low rates of stone tool discard. Bone tools were apparently easier to acquire, and here they may have partly replaced stone tools for some purposes[13].

Comparative analysis of additional sites will be necessary to evaluate the incidence of butchery among preserved late Pleistocene mastodons and the impact of human hunting on mastodon populations[27]. However, the patterns emerging here suggest that mastodon butchery may have been much more common than previously recognized. These observations provide substantial, if indirect, support for the hypothesis that human hunting was an important factor in the ultimate extinction of mastodons.

I thank G. F. Jarvis, W. Egeler, G. Fletcher, R. and P. Van Sickle, C. Austin, D. and S. Johnson, G. and J. Taylor and P. Wesley for their assistance in the discovery of new mastodon sites and in donating material for study. I also thank the following for comments on the manuscript or other assistance: C. E. Badgley, R. Bonnichsen, S. J. Carlson, C. S. Darling-Fisher, J. A. Dorr Jr, D. F. Eschman, W. R. Farrand, R. I. Ford, G. C. Frison, P. D. Gingerich, K. Klitz, J. Krakker, P. S. Martin, R. E. Morlan, W. J. Parry, P. Shipman, G. R. Smith, J. D. Speth, D. Stanford and H. T. Wright. The research was supported by a grant from the Scott Turner Fund, Department of Geological Sciences, University of Michigan.

Received 24 October 1983; accepted 3 January 1984.

1. Mason, R. J. *Curr. Anthrop.* **3,** 227–278 (1962).
2. Müller-Beck, H. *Science* **152,** 1191–1210 (1966).
3. Haynes, C. V. Jr *Science* **166,** 709–715 (1969).
4. Saunders, J. J. *Can. J. Anthrop.* **1,** 87–98 (1980).
5. Irwin-Williams, C. in *Pleistocene Extinctions* (eds Martin, P. S. & Wright, H. E. Jr) 337–347 (Yale University, New Haven, 1967).
6. Bryan, A. L., Casamiquela, R. M., Cruxent, J. M., Gruhn, R. & Ochsenius, C. *Science* **200,** 1275–1277 (1978).
7. Gustafson, C. E., Daugherty, R. & Gilbow, D. W. *Can. J. Archaeol.* **3,** 157–164 (1979).
8. Graham, R. W., Haynes, C. V. Jr, Johnson, D. L. & Kay, M. *Science* **213,** 1115–1117 (1981).
9. Fisher, D. C. & Koch, P. L. *Geological Society of America, Abstracts with Programs* **15,** 573 (1983).
10. Bonnichsen, R. *Archaeol. Surv. Can.* **89,** (1979); *Can. J. Anthrop.* **2,** 137–144 (1982).
11. Morlan, R. E. in *Pre-Llano Cultures of the Americas: Paradoxes and Possibilities* (eds Humphrey, R. L. & Stanford, D.) 125–145 (Anthrop. Soc., Washington, DC, 1979); *Archaeol. Surv. Can.* **94** (1980).
12. Stanford, D. in *Pre-Llano Cultures of the Americas: Paradoxes and Possibilities* (eds Humphrey, R. L. & Stanford, D.) 101–123 (Anthrop. Soc., Washington, DC, 1979).
13. Stanford, D., Bonnichsen, R. & Morlan, R. E. *Science* **212,** 438–440 (1981).
14. Farrand, W. R. *Anthrop. Pap. Mus. Anthrop. Univ. Mich.* **61,** 74–84 (1977).
15. Shipman, P., Fisher, D. C. & Rose, J. *Paleobiology* (submitted).
16. Potts, R. & Shipman, P. *Nature* **291,** 577–580 (1981).
17. Shipman, P. *Ann. N.Y. Acad. Sci.* **376,** 357–386 (1981).
18. Shipman, P. & Rose, J. *J. anthrop. Archaeol.* **2,** 57–98 (1983).
19. Shipman, P. & Foster, G. *J. archaeol. Sci.* (submitted).
20. Fisher, D. C. *Paleobiology* (submitted).
21. Rancier, J., Haynes, G. & Stanford, D. *Southwestern Lore* **48,** 1–17 (1982).
22. Miller, S. J. & Dort, W. Jr in *Early Man in America from a Circum-Pacific Perspective* (ed. Bryan, A. L.) 129–139 (University of Alberta, Edmonton, 1978).
23. Frison, G. C. *Prehistoric Hunters of the High Plains* (Academic, New York, 1978).
24. Johnson, E. *Can. J. Anthrop.* **2,** 145–157 (1982).
25. Wheat, J. B. *Can. J. Anthrop.* **2,** 169–177 (1982).
26. Fitting, J. E. *Anthrop. Pap. Mus. Anthrop. Univ. Mich.* **27,** 1–81 (1966).
27. Martin, P. S. in *Pleistocene Extinctions: The Search for a Cause* (eds Martin, P. S. & Wright, H. E. Jr) 75–120 (Yale University Press, New Haven, 1967).

Desynchronization of the oral temperature circadian rhythm and intolerance to shift work

A. Reinberg*, P. Andlauer†, J. De Prins‡, W. Malbecq‡, N. Vieux§ & P. Bourdeleau§

* E. R. Chronobiologie Humaine CNRS No. 105, Fondation A. de Rothschild, 29 rue Manin, 75940 Paris Cedex 19, France
† Gillette France, BP, 26, F-74010 Annecy Cedex, France
‡ Departement de Physique, Laboratoire Etalons de Frequence, Université Libre de Bruxelles, 1050 Bruxelles, Belgium
§ Shell Francaise, BP 1, F-76650 Le Petit Couronne, France

The present study tested the hypothesis that subjects with a good tolerance to shift work maintain the circadian period τ of their temperature rhythm equal to 24 h, while τ may differ from 24 h when subjects exhibit one or several clinical signs of intolerance. These latter are mainly: persisting sleep disturbance, persisting fatigue, changes in mood and behaviour, and digestive troubles, from gastritis to overt peptic ulcer[1,2]. These symptoms were used here to classify the subjects studied. Medications, including all types of sleeping pills, are ineffective. As was the case in the present study, some subjects may tolerate shift work for 35 yr, reaching 57 yr of age without complaint, while others, after several months or many years, quite rapidly (within 6 months) develop symptoms of intolerance[1].

Three types of male subjects were involved in the study: (1) Thirty-eight with a good tolerance, aged 21–58 yr (mean age 41.0 yr); duration of shift work 6 months to 36 yr (mean 15.2 yr). (2) Twenty-seven subjects with a poor tolerance, aged 21–58 yr (mean age 41.1 yr); duration of shift work 6 months to 35 yr (mean 18.5 yr). Subjects in these two groups were shift working at the time of the study, while the 18 subjects of the third group had been transferred to positions of diurnal work due to their poor tolerance to shift work over 0.5–4.0 yr (aged 32–58 yr, mean age 46.6 yr; duration of shift work 8–27 yr, mean 16.7 yr). Different types of shift rotation and industry were investigated. (Particulars can be provided on request.)

Subjects volunteered for the study. During a 16–30-day span they measured and recorded their oral temperature at least every 4 h and at most at alternating hours (five to eight times a day, not during sleep) using clinical mercury thermometers (0.05 celsius precision). The actual time point of any measurement (clock hour and minute) was also recorded. They also

© 1984 Nature Publishing Group

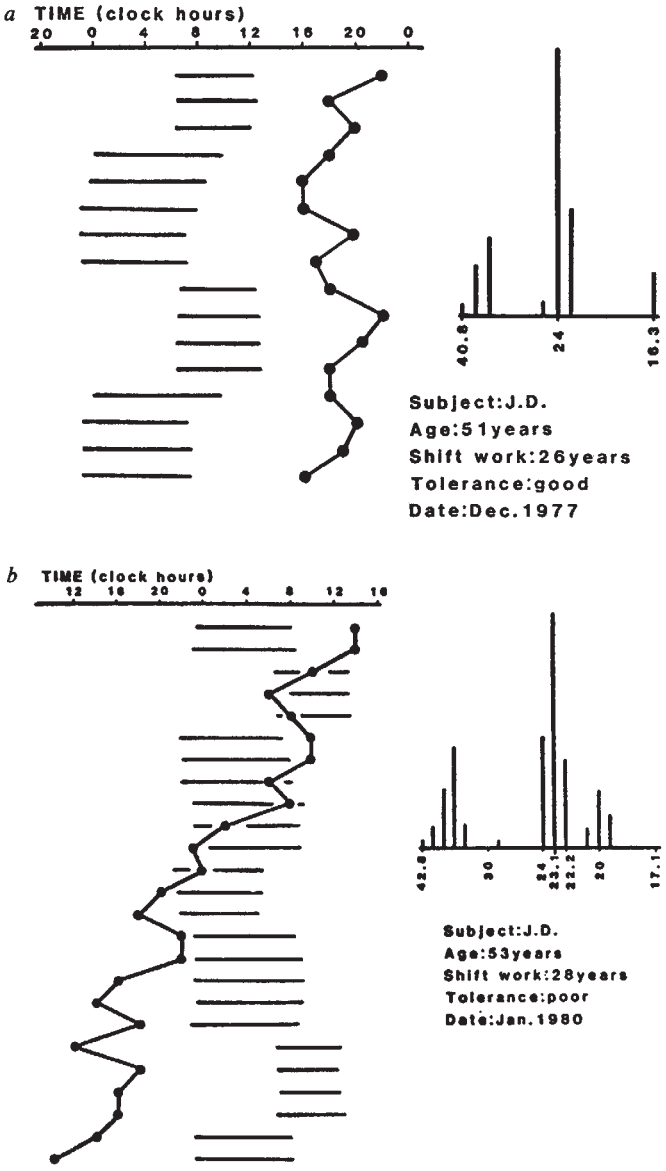


Fig. 1 *a*, Persistence of 24-h periods in the oral temperature circadian rhythm of subject J.D. when his tolerance to shift work was good (December 1977). Left: horizontal bars represent both duration and location of rest (lights-off to lights-on) governed by shift schedules (rapid rotation); black dots and solid line represent day by day acrophase (ϕ) location (peak time) of the temperature rhythm. Right: power spectrum of the temperature rhythm with period τ in hours. Prominent = 24 h. *b*, Desynchronization of the oral temperature circadian rhythm of subject J.D. when his tolerance to shift work was poor (January 1980). Data presentation as in *a*. The oral temperature ϕ is drifting to the left, indicating that $\tau < 24$ h. The power spectrum analysis shows a prominent line for $\tau = 23.1$ h.

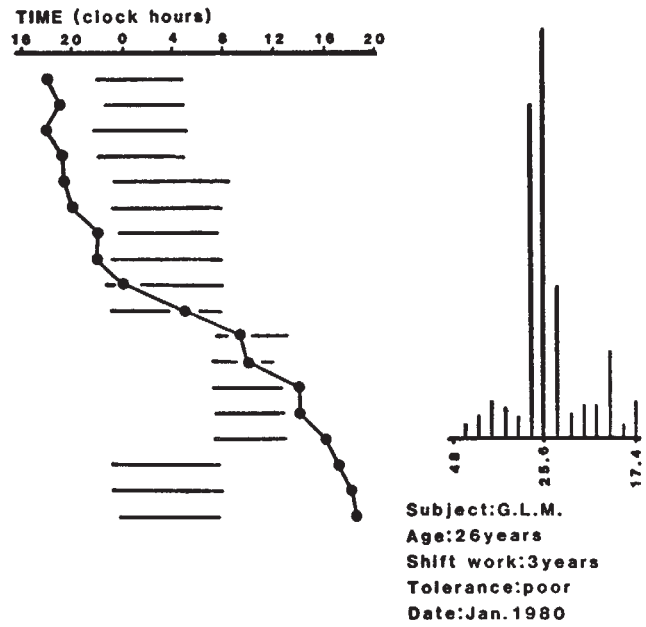


Fig. 2 Desynchronization of the oral temperature circadian rhythm of subject G.L.M. with a poor tolerance to shift work. Data presentation as in Fig. 1*a*. The oral temperature ϕ is drifting to the right, indicating that $\tau > 24$ h. The power spectrum analysis shows a prominent line for $\tau = 25.6$ h.

recorded times of lights-on and lights-off as well as type of shift. During the study, subjects were neither taking medication nor suffering from infectious diseases, as both factors can affect[3] circadian changes of temperature, including τ.

Two major difficulties of the study were that the data were gathered at unequal time intervals and the possibility of masking effects due to lack of temperature measurements during sleep[4]. The mathematical solution to this problem is derived from conventional spectral analyses, that is, the estimation of a correlation function and the calculation of its Fourier transform. The problem of the determination of correlations was resolved using a class estimator which was mathematically validated[5]. Furthermore, the most conventional method, consisting of a day by day display of estimated acrophases (peak time location in the 24-h schedule)[4,6], was also used to visualize any acrophase (peak time ϕ) drift or shift. Each ϕ estimate corresponded to the crest of the best-fitting cosine function approximating all data for 24 h.

These two methods have limitations. For instance, the prominent amplitude of τ on the power spectrum was depicted on the graph printed by the computer, without statistical significance. The display of estimated acrophases was useful for visualizing ϕ drift or shift but not for quantifying it. Therefore, it was of interest to use both of these methods to test the concordance of results provided, as can be seen in the figures.

Thirty-one out of 38 subjects with good or excellent tolerance to shift work showed a prominent circadian period τ equal to 24 h. In contrast, only 5 out of 27 subjects with poor tolerance to shift work had a prominent τ of 24 h. The χ^2 test between these two groups was 25.37 with $P < 0$. Figures 1–3 show the ϕ values and the power spectrum of τ values for subjects with good and poor tolerance. We were able to study one subject while he was still tolerating shift work (Fig. 1*a*) and again 2 yr later when he had become intolerant (Fig. 1*b*). It seems that the temperature rhythm was free-running in subjects with poor tolerance and that τ was either shorter (21.4 ± 0.52 h, s.e.m., mean τ of 10 cases) or longer (25.76 ± 0.25 h, mean of 8 cases).

Sixteen out of 18 subjects who suffered from signs of poor tolerance and who during the previous 0.5–4 yr had resumed exclusively diurnal work had a prominent τ equal to 24 h. At the time of their individual study, at most only minor complaints were reported. Subject S.A. was studied when shift working and suffering poor tolerance and had a prominent τ of 26.6 h (Fig. 3*a*). He was studied again about 1 yr after his transfer to diurnal work, and the prominent τ was then 24 h (Fig. 3*b*).

It seems, therefore, that some subjects may exhibit a free-running circadian rhythm of temperature after several months or many years of shift work. Despite the shifts, the activity–rest rhythm of these subjects showed a mean of 24 h. In other words, a desynchronization with differences in τ values between two rhythms (activity–rest and temperature) occurred in a majority of subjects with poor tolerance. Similar desynchronizations have already been reported[4,6,7] at least for some subjects, during isolation experiments performed without time clue and cue. It seems also that desynchronization of rhythm found in most

© 1984 Nature Publishing Group

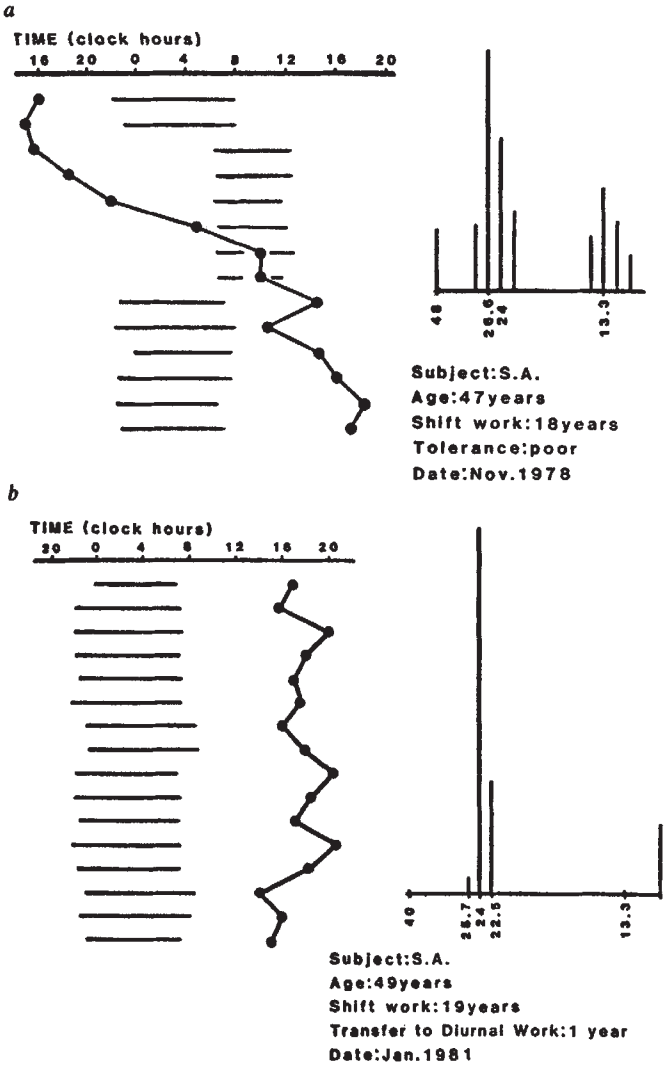


Fig. 3 *a*, Desynchronization of the oral temperature rhythm of subject S.A. with poor tolerance to shift work (November 1978). Data presentation as in Fig. 1*a*. Prominent line for $\tau = 26.6$ h. *b*, 24 h synchronized rhythm of subject S.A. about 1 yr after his transfer to diurnal work (January 1981). Data presentation as in preceding figures.

subjects with poor tolerance was not related to the subject's age, duration of shift work, type of shift rotation, or type of industry because there was no major difference with regard to any of these factors among groups with poor and good tolerance.

We thank the managements and personnel of Shell Francaise, Le Petit Couronne and Gillette-France, Annecy, for their help and technical assistance. Support was also provided by CNAMTS-France and by the US Army European Research Office, ARI Behavioral and Social Sciences L.O., contract ERO 134-82. Special thanks are due to the 83 subjects who volunteered for the study and to Drs G. Duverneuil, A. Robert, A. Pommier, P. Quelin and J. Rety for their medical cooperation.

Received 11 July; accepted 20 December 1983.

1. Andlauer, P., Reinberg, A., Fourre, L., Battle, W. & Duverneuil, *J. Physiol., Paris* **75**, 507 (1979).
2. Koller, M. *et al.* in *Night and Shift work: Biological and Social Aspects* (eds Reinberg, A., Vieux, N. & Andlauer, P.) 465–472 (Pergamon, Oxford, 1981).
3. Reinberg, A. & Smolensky, M. in *Body Temperature: Control in Health and Diseases and the Effects of Drugs* (eds Lomax, P. & Schonbaum, E.) *Pharmac. Ther.* **22**, 425–464 (Pergamon, Oxford, 1983).
4. Wever, R. A. *The Circadian System of Man. Results of Experiments under Temporal Isolation* (Springer, New York, 1979).
5. De Prins, J. & Malbecq, W. *Bull. Cl. Sci.* **69**, 287–294 (1983).
6. Aschoff, J. & Wever, R. A. *Fedn Proc.* **35**, 2326 (1976).
7. Weitzman, E. D., Czeisler, C. A. & Moore-Ede, M. C. in *Advances in Sleep Research* Vol. 7 (eds Johnson, L. C., Tepas, D. I., Colquhoun, W. P. & Colligan, M. J.) 75–91 (S. P. Medical and Scientific, New York, 1981).

Proliferation of mature oligodendrocytes after trauma to the central nervous system

Samuel K. Ludwin

Department of Pathology (Neuropathology), Queen's University, Kingston, Ontario, Canada K7L 3N6

It has long been thought that mature oligodendrocytes in the adult mammalian central nervous system (CNS) are post-mitotic and are unable to proliferate in response to injury. The implications of this have been profound, because it has been suggested that this failure of oligodendrocytes to undergo mitosis is perhaps one of the reasons for the failure of the human CNS to undergo remyelination after demyelinating disease. This is in contrast with the normal peripheral nervous system in which there is consistent remyelination, and brisk Schwann cell mitosis. Although it has recently been shown that oligodendrocytes can be regenerated following some specific instances of demyelination[1–3], it has long been accepted that unlike mature astrocytes and microglia (macrophages), oligodendrocytes do not proliferate in response to general conditions damaging the nervous system[4,5]. Here we show that mature oligodendrocytes in adult animals, as well as astrocytes and microglia, are able to respond to damage in the CNS following trauma by incorporating tritiated thymidine into their nuclei.

Adult (3–4 month old) male mice of the CD1 strain (Charles River, Canada) were briefly anaesthetized with methoxyflurane, and a small bone flap over the right parieto-occipital cortex was removed. A small piece of cortex, measuring ~3 mm, was scooped out shallowly from the dorsal cortex. The wound was packed with Gelfoam, and the skin sutured. Two days later, the animals were injected intraperitoneally with a single dose of tritiated thymidine, 6 μCi per g body weight. After 24 h the animals were perfused with Karnovsky's solution, and the brains sectioned. The brains were sliced coronally. The posterior part of the brain was embedded in paraffin, and coronal sections of the whole brain slice prepared for light microscopy. White matter and cortex, both on the side of the lesions and on the opposite side were taken from the anterior part of the brain and processed for electron microscopy. Both the light microscopic and the electron microscopic sections were prepared for autoradiography as described previously[2]. Control unoperated animals were injected with tritiated thymidine 24 h before killing and the brains processed in the same way.

In control animals, examination of light microscopic autoradiographs revealed almost no mitotic activity within the cortex or white matter after the short 24-h exposure. The subependymal zone containing the germinal matrix cells showed numerous labelled nuclei. Examination of the brains in the operated animals revealed a necrotic cortical defect dorsally on the right side, occasionally involving the underlying white matter. In contrast to the controls, the brains of operated animals showed large numbers of cells whose nuclei were covered with silver grains. Around the area of the wound and in the wound itself numerous labelled nuclei, many of which were macrophages, were seen. Endothelial cells were heavily labelled. The adjacent areas of cortex, in which there was no obvious morphological damage, showed numerous glial nuclei containing grains; the identity of these cells could not always be determined at the light microscopic level. The number of labelled cells in the uninvolved cortex decreased further away from the wound, but labelled cells were seen on the lateral surface of the operated hemisphere, and were found in small numbers in the cortex on the opposite side. Labelled cells were very numerous in the basal ganglia and hippocampus beneath the white matter. Within

© 1984 Nature Publishing Group

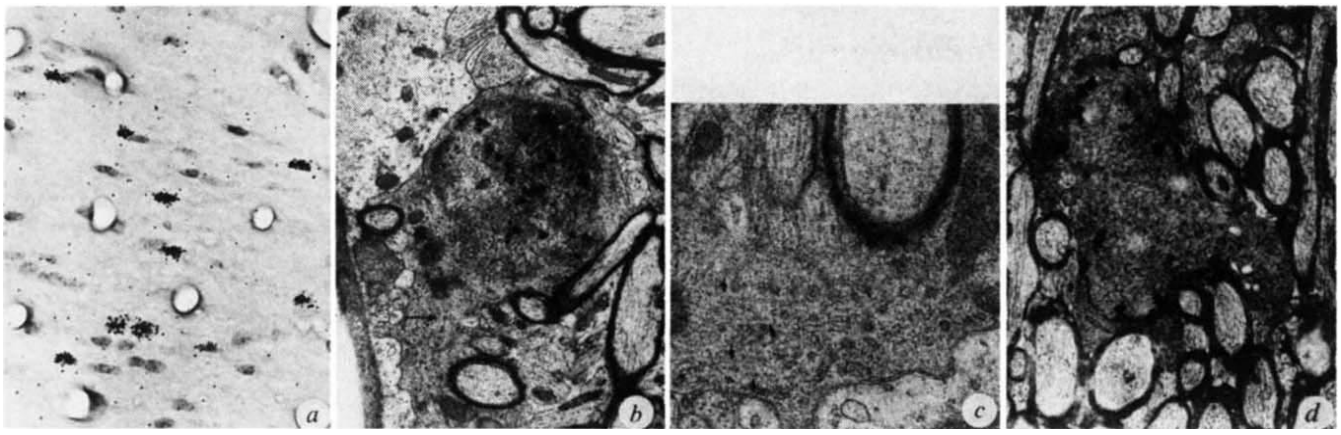


Fig. 1 *a*, Corpus callosum from a mouse 3 days after craniotomy. The autoradiograph reveals numerous labelled nuclei, some elongated and spindle-shaped suggesting intrafascicular oligodendrocytes, others larger and with a rounded shape suggesting astrocytes. Haematoxylin and eosin counter-stain. ×96. *b*, Electron micrograph showing an oligodendrocyte with a labelled nucleus from the deep cortex 3 days after craniotomy, taken from an area away from the wound. An astrocyte is present at upper left and a blood vessel at lower left. The area of cytoplasm marked by an arrow is shown at higher magnification in *c*. ×12,500. *c*, Higher magnification of the oligodendrocyte in *b* showing typical cytoplasmic features. Numerous microtubules (arrows) run parallel to the cytoplasmic process. ×18,600. *d*, Intrafascicular oligodendrocyte from the corpus callosum away from the lesion. Numerous silver grains are seen overlying the nucleus. ultrastructural autoradiograph. ×5,500.

the subcortical white matter and the corpus callosum, numerous cells having the morphology of astrocytes and of intrafascicular oligodendrocytes (Fig. 1*a*) could be readily seen. These labelled cells were found in the absence of any obvious direct damage to the white matter and corpus callosum. The labelled cells were found in the white matter of the corpus callosum extending well across the midline to the opposite hemisphere.

At the electron microscopic level, the labelled cells were identified as astrocytes, macrophages and oligodendrocytes. The macrophages were most prominent in the area around the cortical wound. The ultrastructural study of the oligodendrocytes revealed the features of medium to dark oligodendrocytes with typical round to oval dark nuclei containing clumped nuclear chromatin, dark cytoplasm containing many ribosomes, prominent Golgi apparatus and microtubules (Fig. 1*b*, *c*). The dense bodies, lysosomes and long stringy endoplasmic reticulum characteristic of microglia were not present in these cells. Intrafascicular oligodendrocytes lying between myelinated axons in uninjured white matter and displaying the classical features of medium to dark oligodendrocytes were also seen to be labelled (Fig. 1*d*).

Oligodendroglial proliferation in normal immature animals has been well established and labelled cells are easily demonstrable[4]. In adult mice, after 24 h of exposure to tritiated thymidine, however, almost no labelled oligodendrocytes are found. The experiment described here demonstrates clearly that the mature oligodendrocyte in the adult mammalian brain, like the astrocyte, can respond to simple CNS damage by taking up tritiated thymidine, which is an accepted indication of cell mitosis. This is in contrast to previous studies on the glial response to trauma which have demonstrated mitotic response as seen by incorporation of tritiated thymidine into astrocytes, macrophages and monocytes, but not into oligodendrocytes[5,6]. Another study has shown that oligodendrocytes in the dentate gyrus take up tritiated thymidine after ventral hippocampal commissurotomy[7], but this study was performed by light microscopy and no ultrastructural confirmation of the oligodendrocytic nature of these cells was made. The time involved in the present experiment—24 h—suggests that the medium to dark cells that took up thymidine were of a relatively mature variety to begin with, as it has been shown in normal development that oligodendrocyte maturation requires a few days at each level during the cell's progression from light through to medium and dark cells[8,9]. The distribution of the labelled responsive cells around the lesion showed a gradual decrease in numbers away from the wound in the cortex, but in the white matter and corpus callosum, labelled cells could be seen across the midline to the opposite hemisphere. This distribution has some similarities to the local spread of oedema following a wound, and although the nature of the stimulus to proliferate is unknown it is possible that whatever factors are present may be spread together with the spread of oedema fluid. Another possibility is that these cells are responding to axonal reactions following neuronal death which, although not seen morphologically, may be present at a molecular or chemical level. A recent study[10], however, has failed to show oligodendrocyte proliferation following wallerian degeneration.

In the mutant mouse *Jimpy*, the hypomyelination is accompanied by increased oligodendrocyte proliferation and a shortened lifespan[11,12], but this is a genetic condition and probably has little to do with normal glial responses. The present experiments, however, show incorporation of thymidine (a well accepted marker of cell division) into oligodendrocytes, as a response to a general nonspecific stimulus similar to that seen in other glial cells.

I thank Ms Mirta Chiong for technical assistance and Mrs P. Scilley for typing the manuscript. This work was supported by MRC grant MA. 5818.

Received 5 December; accepted 16 December 1983.

1. Herndon, R. M., Price, D. L. & Weiner, L. P. *Science* **195**, 693–694 (1977).
2. Ludwin, S. K. *Am. J. Path.* **95**, 683–690 (1979).
3. Johnson, E. S. & Ludwin, S. K. *Acta neuropath.* **53**, 93–98 (1981).
4. Mori, S. & Leblond, C. P. *J. comp. Neurol.* **139**, 1–30 (1970).
5. Adrian, E. K. Jr, Williams, M. G. & George, F. C. *J. comp. Neurol.* **180**, 815–840 (1978).
6. Kitamura, T., Tsuchihashi, Y. & Fujita, S. *Acta neuropath.* **44**, 31–39 (1978).
7. Avenda, C. & Cowan, W. M. *Anat. Embryol.* **157**, 347–366 (1979).
8. Imamoto, K., Paterson, J. A. & Leblond, C. P. *J. comp. Neurol.* **180**, 115–128 (1978).
9. Paterson, J. A. *Anat. Anz.* **153**, 149–168 (1983).
10. Fulcrand, J., Valat, J. & Privat, A. *Reprod. Nutr. Dev.* **22**, 171–178 (1982).
11. Privat, A., Valat, J., Lachapelle, F., Baumann, N. & Fulcrand, J. *Brain Res.* **254**, 411–416 (1981).
12. Skoff, R. P. *Brain Res.* **248**, 19–31 (1982).

© 1984 Nature Publishing Group

Central target for the behavioural effects of vasopressin neuropeptides

David de Wied, Odile Gaffori, Jan M. van Ree & Wybren de Jong

Rudolf Magnus Institute for Pharmacology, Medical Faculty, University of Utrecht, Vondellaan 6, 3521 GD Utrecht, The Netherlands

The neurohypophysial hormone vasopressin exerts antidiuretic, vasopressor and behavioural effects (for example, facilitation of memory processes)[1–6]. Vasopressin may alter animal behaviour via direct effect on brain processes[2,3,6]. Recently, however, it has been suggested that vasopressin acts mainly at peripheral receptor systems and influences behavioural mechanisms by altering visceral afferent signals[7]. We now present data showing that (1) central administration of [Arg[8]]vasopressin (AVP) and more potently [pGlu[4], Cyt[6]]AVP(4–8), the desglycinamide derivative of a peptide generated from AVP by brain synaptic membranes[8], produce the behavioural effect (promotion of passive avoidance behaviour) without the pressor effect; (2) central administration of a vasopressor antagonist blocks the behavioural but not the pressor effect of systemically administered AVP; and (3) [pGlu[4], Cyt[6]]AVP(4–8) induces the behavioural effect in the absence of the pressor effect. The results indicate that AVP and related peptides affect passive avoidance behaviour by a direct central action and that the structural requirement for activation of central vasopressin receptors differs from that of the peripheral cardiovascular receptors, although both can be blocked by the same vasopressor antagonist.

For behavioural experiments rats (140–160 g) of an inbred strain were trained in a passive avoidance test procedure[9]. Rats were placed on an illuminated platform, and allowed to enter a large dark compartment equipped with a grid floor. On the next day three more trials an unavoidable scrambled footshock (0.25 mA, 2 s) was delivered through the grid floor of the dark compartment (learning trial). Passive avoidance latencies were determined at 24 h after the learning trial, by placing the rat on the platform and measuring the latency (up to a maximum of 300 s) to enter the dark compartment. The rats were treated immediately after the learning trial. For intracerebroventricular injection the rats were equipped with a polyethylene cannula 1 week before behavioural testing. At the end of the experiment correct localization of each cannula was ascertained[6].

Subcutaneous (s.c.) injection of 3.0 μg [Arg[8]]vasopressin (AVP) and 0.3 μg [pGlu[4], Cyt[6]]AVP(4–8) (in this paper referred to as AVP(4–8)) increased passive avoidance latencies (Table 1, Fig. 1). A similar effect was observed when lower amounts of these peptides were injected intracerebroventricularly (i.c.v.) (Table 1, Fig. 1). Neither AVP nor AVP(4–8) affected the latencies in non-shocked animals. To determine an ED_{50} for this effect, groups of at least six rats were injected either s.c. or i.c.v. with graded doses of these peptides (AVP s.c.: 0.1, 0.3, 1.0 μg; i.c.v.: 0.1, 0.3, 3.0 ng; AVP(4–8) s.c.: 0.03, 0.1, 0.3 μg; i.c.v.: 0.3, 3.0, 30 pg). Lower and higher doses were ineffective or induced maximal effects respectively. The ED_{50} was calculated by the method of Litchfield and Wilcoxon[10]; a latency of more than 190 s was taken to be a response (only one of 97 placebo-treated rats scored positive). The ED_{50} for AVP appeared to be 0.57 μg and 1.3 ng after s.c. and i.c.v. treatment respectively and for AVP(4–8) 0.105 μg and 4.5 pg. Thus, on a molar base AVP(4–8) is approximately 4 and 200 times more potent than AVP in promoting passive avoidance behaviour, when s.c. and i.c.v. treatments are compared.

Low doses (0.3 and 1 μg s.c.) of the vasopressor antagonist $d(CH_2)_5$Tyr(Me)AVP (ref. 11), in the absence of agonist, attenuated passive avoidance behaviour. A somewhat higher dose (3 μg), however, did not affect the behavioural response (Table 1). This dose was, therefore, selected to investigate the interaction of this peptide with the behavioural effects of AVP and AVP(4–8). The vasopressor antagonist injected s.c. immediately after the learning trial and 30 min before s.c. treatment with AVP or AVP (4–8), antagonized the action of these peptides (Table 1, Fig. 1*a*). Dose–response curves were assessed with AVP(4–8) in the presence of two dose levels of the antagonist. The data revealed that the antagonism between these peptides was competitive in nature (Fig. 1*a*). The same was found after i.c.v. injection of the antagonist and AVP or AVP(4–8) (Fig. 1*b*). The dose–response curve of both peptides shifted to the right in the presence of the antagonist. The slope of the curves obtained in the presence of the antagonist was somewhat steeper than in the absence of the antagonist. These data suggest that the vasopressor antagonist acts directly on the receptor sites that are activated by AVP and AVP(4–8) and that mediate the effect of these peptides on passive avoidance behaviour.

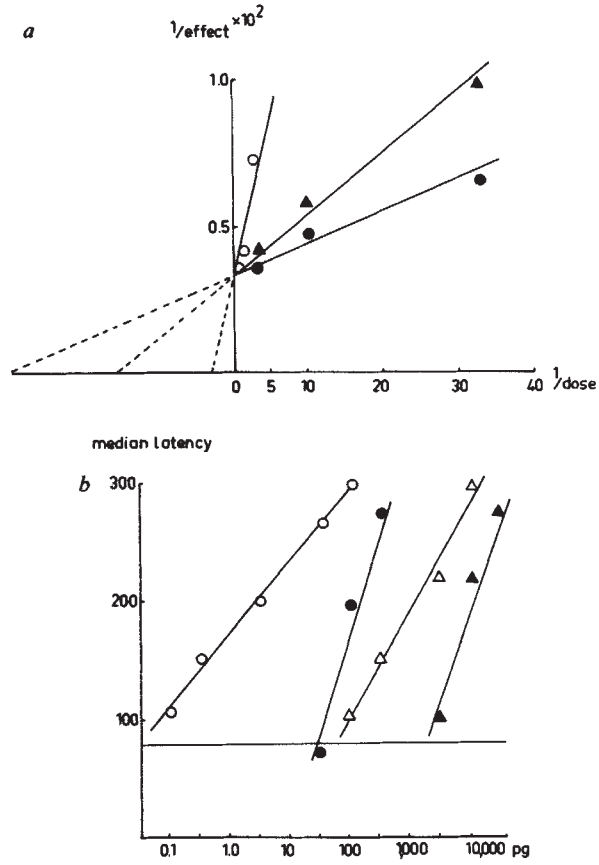


Fig. 1 Antagonism between the vasopressor antagonist, $d(CH_2)_5$Tyr(Me)AVP and [Arg[8]]vasopressin (AVP) or [pGlu[4]-Cyt[6]]AVP(4–8) (AVP(4–8)) following s.c. (*a*) or i.c.v. (*b*) treatment using passive avoidance behaviour as test parameter. The antagonist was injected immediately after the learning trial and the agonists 30 min later. Retention of the response was determined by timing the latency to enter the dark compartment 24 h after the learning trial. *a*, Groups of rats were s.c.-treated with graded doses of AVP(4–8) after injection of placebo (●—●) or the antagonist (▲—▲ 3 μg, s.c. or ○—○ 10 μg, s.c.). Graphs were constructed according to Lineweaver–Burk. The reciprocal value of the dose (μg) is plotted against the reciprocal value of the median of the measured latencies (s). Theoretically, competitive antagonism is present when the common intersection of the calculated lines is at the *y*-axis (as in the graph), whereas in case of non-competitive antagonism the common intersection of the extended lines (dotted lines) is at the *x*-axis. *b*, Groups of rats were i.c.v.-treated with graded doses of AVP (△—△) or AVP(4–8) (○—○) after i.c.v. injection of placebo (open symbols) or 3 ng of the antagonist (filled symbols). The dose (pg) of AVP and AVP(4–8) is plotted against the median of the measured latencies (s). Lines were calculated with the methods of least squares.

© 1984 Nature Publishing Group

Table 1 Interaction between a vasopressor antagonist and AVP or AVP(4–8) following s.c. or i.c.v. treatment using the passive avoidance response as test parameter

Treatment*		Latency (s)	
Placebo s.c.	Placebo s.c.	101	(51–162)†
Antagonist 3 μg s.c.	Placebo s.c.	94	(42–127)
Placebo s.c.	AVP 3 μg s.c.	267‡	(163–300)
Antagonist 3 μg s.c.	AVP 3 μg s.c.	99‖	(18–127)
Placebo s.c.	Placebo i.c.v.	77	(24–112)
Antagonist 3 μg s.c.	Placebo i.c.v.	81	(21–110)
Placebo s.c.	AVP 3 ng i.c.v.	280†	(64–300)
Antagonist 3 μg s.c.	AVP 3 ng i.c.v.	57‖	(11–69)
Placebo i.c.v.	Placebo s.c.	82	(56–93)
Antagonist 3 ng i.c.v.	Placebo s.c.	95	(82–131)
Placebo i.c.v.	AVP 3 μg s.c.	300‡	(183–300)
Antagonist 3 ng i.c.v.	AVP 3 μg s.c.	69‖	(54–103)
Placebo i.c.v.	Placebo s.c.	77	(24–112)
Antagonist 3 ng i.c.v.	Placebo s.c.	70	(13–110)
Placebo i.c.v.	AVP(4–8) 0.3 μg s.c.	275‡	(76–300)
Antagonist 3 ng i.c.v.	AVP(4–8) 0.3 μg s.c.	99‖	(22–172)

The vasopressor antagonist used was d(CH_2)Tyr(Me)AVP. Peptides were injected immediately (antagonist) and 30 min after the learning trial and the latency to enter the dark compartment was measured 24 h later.

* Doses refer to amount of peptide per rat.

† Medium within brackets the 25th and 75th percentile of six rats per treatment.

‡ $P<0.02$. Different as compared with simultaneous treatment with placebo instead of AVP or AVP(4–8) (Mann–Whitney *U*-test).

§ $P<0.05$; ‖ $P<0.02$. Different as compared with simultaneous treatment with placebo instead of antagonist (Mann–Whitney *U*-test).

Table 2 Interaction between a vasopressor antagonist and AVP or AVP(4–8) following s.c. and i.c.v. treatment using changes in blood pressure and heart rate as test parameters

Treatment	Blood pressure* Δ mm Hg	Heart rate* Δ b.p.m.	No. of rats
Placebo (s.c.)	−0.6±2.6	5±10	13
AVP (3 μg, s.c.)	32.3±3.3‡	−112±19‡	6
AVP(4–8) (3 μg, s.c.)	2.0±1.5	−9±9	8
Placebo (i.c.v.)	2.8±2.3	−3±9	10
AVP (30 ng, i.c.v.)	−1.8±3.7	−17±9	4
AVP(4–8) (0.3 ng, i.c.v.)	−3.2±5.6	17±15	4
Antagonist (3 ng, i.c.v.)	−0.7±4.1	31±6	4
Placebo (s.c.) + placebo (s.c.)	−2.4±2.7	18±6	4
Antagonist (3 μg, s.c.) + placebo (s.c.)	3.2±1.2	16±7	5
Placebo (s.c.) + AVP (3 μg, s.c.)	41.2±4.5‡	−100±14‡	5
Antagonist (3 μg, s.c.) + AVP (3 μg, s.c.)	0.0±2.5§	−32±6§	5
Placebo (i.c.v.) + AVP (3 μg, s.c.)	36.0±3.0	−137±18	5
Antagonist (3 ng, i.c.v.) + AVP (3 μg, s.c.)	34.2±6.5	−140±8	5

* Data are expressed as mean (±s.e.m.) changes in blood pressure (Δ mm Hg) and heart rate (Δ beats per min) at 30 min after last peptide injection as compared with baseline values, obtained before treatment. Baseline values of all rats: blood pressure 113.2±2.0 mm Hg; heart rate: 386±7 b.p.m. The various groups did not differ with respect to baseline values.

† Doses are amount of peptide per rat. The two treatments were given with a time interval of 30 min.

‡ Different for placebo-treated rats ($P<0.001$, Student's *t*-test).

§ Different from placebo- instead of antagonist-treated rats ($P<0.005$, Student's *t*-test).

The potency difference between s.c. and i.c.v. treatment with AVP and AVP(4–8) (approximately 440 and 23,000 times for AVP and AVP(4–8) respectively) demonstrates that these receptor sites are located in the brain. This suggestion is further supported by data of additional experiments. When the antagonist was given s.c. and AVP was injected i.c.v. in a dose of 3 ng which is inactive after systemic treatment, the effects of AVP on passive avoidance behaviour were completely antagonized by the vasopressor antagonist (Table 1). Furthermore, a similar antagonism was found when the antagonist was injected i.c.v. and AVP or AVP(4–8) was administered s.c. (Table 1). These data strongly suggest that the antagonism between the vasopressor antagonist and AVP or AVP(4–8) occurs in the central nervous system both following s.c. and i.c.v. treatment of the peptides.

Next, the influence of these peptides on blood pressure was determined. Rats (200–220 g) were equipped with indwelling cannulae in the ventral tail artery 1 day before blood pressure registration was performed. The conscious rats were tested in their home cage and effects of s.c. or i.c.v. treatment were assessed for at least 1 h. Subcutaneous administration of AVP (0.3–3.0 μg) induced a dose-dependent increase in blood pressure associated with a bradycardia. The maximal effect was obtained between 20 and 40 min after treatment. Data of the highest dose at 30 min after injection are presented in Table 2. In contrast, s.c. injection of identical doses of AVP(4–8) did not affect blood pressure and heart rate at any time after injection (Table 2). Intraventricular injection of 30 ng AVP or 0.3 ng AVP(4–8), which exert maximal behavioural effects, did not induce a pressor response or bradycardia. The vasopressor antagonist (3.0 μg s.c.) injected 30 min before AVP did not affect blood pressure or heart rate *per se*, but prevented AVP-induced pressor response and bradycardia. Intracerebroventricular administration of the antagonist (3 ng) which can block the behavioural action of s.c.-i.c.v.-injected AVP and AVP-(4–8) (Table 1, Fig. 1) had no effect on the pressor response of s.c.-injected AVP (3 μg) (Table 2). These data suggest that AVP exerts its action on blood pressure and heart rate by activation of peripherally located receptor systems, which can be blocked by the vasopressor antagonist. These receptor systems apparently cannot be activated by AVP(4–8) and appear not to be involved in mediating the effect of the peptides on passive avoidance behaviour.

The present data indicate that the vasopressin receptors mediating the pressor response are located in the periphery, while those mediating the behavioural effects are probably present in the central nervous system. While the 'pressure' receptor can hardly be activated by AVP(4–8), this peptide can potently activate the 'behavioural' receptor. The potent AVP(4–8) peptide, derived from [pGlu[4], Cyt[6]]AVP(7–9), that is generated from vasopressin[6] and present in the brain (unpublished data), may be related to important endogenous molecules for modulation of behavioural processes. Interestingly, both the pressor and behavioural receptors can be blocked by the vasopressor antagonist used in the present studies. The findings extend previous observations[12–14] and support the concept that the main locus of the behavioural action of vasopressin and related neuropeptides is in the brain[4–6,15].

We thank Dr M. Manning for supplying the vasopressor antagonist and Organon International for the other peptides. The technical assistance of Mrs Joke Cox-van Put is acknowledged.

© 1984 Nature Publishing Group

Received 5 September; accepted 30 December 1983.

1. De Wied, D. *Nature* **232**, 58–60 (1971).
2. Van Ree, J. M., Bohus, B., Versteeg, D. H. G. & De Wied, D. *Biochem. Pharmac.* **27**, 1793–1800 (1978).
3. Van Wimersma Greidanus, Tj. B., Van Ree, J. M. & De Wied, D. *Pharmac. Ther.* **20**, 437–458 (1983).
4. De Wied, D. *Life Sci.* **20**, 195–204 (1977).
5. De Wied, D. & Van Ree, J. M. *Life Sci.* **31**, 709–719 (1982).
6. De Wied, D. *Life Sci.* **19**, 685–690 (1976).
7. Le Moal, M. *et al. Nature* **291**, 491–493 (1981).
8. Burbach, J. P. H., Kovács, G. L., De Wied, D., Van Nispen, J. W. & Greven, H. M. *Science* **221**, 1310–1312 (1983).
9. Ader, R., Weijnen, J. A. W. M. & Moleman, P. *Psychon. Sci.* **26**, 125–128 (1972).
10. Litchfield, J. T. Jr. & Wilcoxon, F. *J. pharmac. expl Ther.* **96**, 99–113 (1949).
11. Kruszynski, M. *et al. J. med. Chem.* **23**, 364–368 (1980).
12. Van Wimersma Greidanus, Jj. B., Dogterom, J. & De Wied, D. *Life Sci.* **16**, 637–644 (1975).
13. Kovács, G. L., Bohus, B., Versteeg, D. H. G., De Kloet, E. R. & De Wied, D. *Brain Res.* **175**, 303–314 (1979).
14. De Wied, D., Greven, H. M., Lande, S. & Witter, A. *Br. J. Pharmac.* **45**, 118–122 (1982).
15. Södersten, P., Henning, M., Melin, P. & Ludin, S. *Nature* **301**, 608–610 (1983).

κ- and δ-opioid receptor agonists differentially inhibit striatal dopamine and acetylcholine release

Arie H. Mulder, George Wardeh, Francois Hogenboom & Abraham L. Frankhuyzen

Department of Pharmacology, Free University, Medical Faculty, Van der Boechorststraat 7, 1081 BT Amsterdam, The Netherlands

At least three different families of endogenous opioid peptides, the enkephalins, endorphins and dynorphins, are present in the mammalian central nervous system (CNS). Immunocytochemical studies have demonstrated their localization in neurones[1–8], which supports the view that these peptides may have a role as neurotransmitters or neuromodulators. However, the target cells and cellular processes acted upon by the opioid peptides are still largely unknown. One possible function of neuropeptides, including the opioid peptides, may be presynaptic modulation of neurotransmission in certain neuronal pathways, for example, by inhibition or promotion of neurotransmitter release from the nerve terminals[9–12]. Here we report that dynorphin and some benzomorphans potently and selectively inhibit the release of (radiolabelled) dopamine from slices of rat corpus striatum, by activating κ-opioid receptors. In contrast, [Leu^5]enkephalin and [D-Ala^2, D-Leu^5]enkephalin selectively inhibit acetylcholine release by activating δ-opioid receptors.

It has recently become clear that opioid receptors do not constitute a single population. Analysis of the activity of a variety of opioid compounds in a number of smooth muscle bioassays as well as *in vivo* tests has led to a subdivision of opioid receptors in μ-, δ- and κ-receptors[13–17]. Whereas morphine primarily activates μ-receptors, [Leu^5]enkephalin and [D-Ala^2, D-Leu^5] enkephalin (DADLE) appear to be selective δ-receptor agonists and benzomorphans, such as ketocyclazocine, as well as dynorphin are thought to activate selectively κ-receptors[15–19].

The corpus striatum, a brain structure of major importance for the control of motor activity, receives a rich innervation by dopaminergic neurones emanating from the substantia nigra[20]. The striatum is also abundant in acetylcholine and there is ample proof that striatal cholinergic interneurones are subject to inhibitory control by dopaminergic neurones[20–25]. High levels of the enkephalins are found in the corpus striatum[26–28] and more recently also the presence of dynorphin in this brain region has been demonstrated[29,30]. The primary aim of the present study was to examine whether the release of dopamine and acetylcholine is modulated, possibly in a differential manner, by [Leu^5] enkephalin and dynorphin.

Release of the radiolabelled neurotransmitters was studied using *in vitro* methods[24,31–33]. Small slices prepared from the corpus striatum were incubated in a physiological medium containing low concentrations of ^{3}H-dopamine and ^{14}C-choline to label the neurotransmitter stores in dopaminergic and cholinergic nerve terminals, respectively. Then the slices were superfused with physiological medium and after 40 min depolarization-induced calcium-dependent release of ^{3}H-dopamine and ^{14}C-acetylcholine was induced by exposing the tissue to an elevated (15 mM) K^+ concentration, either in the absence (controls) or presence of the opioid peptides or drugs examined (see legends to tables and figures for experimental details).

Table 1 Effects of dynorphin(1–17) and dynorphin(1–13) on the K^+-induced release of ^{3}H-dopamine and ^{14}C-acetylcholine from rat striatal slices

In presence of	Release (as % of controls) of: ^{3}H-dopamine	^{14}C-acetylcholine
Dynorphin(1–17)		
0.01 μM	78.4 ± 2.0*	102.2 ± 2.5
0.1 μM	68.2 ± 1.8*	99.4 ± 2.3
Dynorphin(1–13)		
0.1 μM	71.4 ± 1.9*	92.5 ± 2.8
1.0 μM	70.2 ± 2.7*	79.8 ± 3.6†

Male Wistar rats (140–160 g body weight) were decapitated, the corpora striata were dissected and the tissue was chopped into small slices (~2×0.3×0.3 mm). The slices were preincubated for 10 min at 37 °C in 5 ml of Krebs–Ringer–bicarbonate medium in a Dubnoff metabolic shaking incubator under a 95% O_2–5% CO_2 atmosphere. The medium contained 121 mM NaCl, 1.87 mM KCl, 1.17 mM KH_2PO_4, 1.17 mM Mg SO_4, 1.22 mM $CaCl_2$, 25 mM $NaHCO_3$ and 11.1 mM glucose; *p*H 7.4. Then incubation was continued for 15 min with medium containing 5 μCi ^{3}H-dopamine (specific activity 47 Ci mmol^{-1}) and 2 μCi ^{14}C-choline (specific activity 59 mCi mmol^{-1}). Subsequently, labelled slices were transferred to small superfusion chambers (0.25 ml volume; about 5 mg wet weight of tissue per chamber), maintained at 37 °C and superfused at a rate of 0.25 ml min^{-1} with medium containing 20 μM bacitracin. In each experiment 16–24 superfusion chambers were run simultaneously. After 30 min of superfusion (that is, $t = 30$ min) four successive 10-min samples were collected. At $t = 40$ min the slices were exposed for 2 min to medium containing 15 mM K^+ (isomolar replacement of NaCl by KCl) and then superfusion was continued with normal medium. Peptides were present in the medium from 10 min before K^+ stimulation throughout the experiment. At the end of the experiment radioactivity remaining in the tissue was extracted with 0.1 M HCl. The radioactivity in superfusion samples and tissue extracts was assayed by liquid scintillation counting. The amount of radioactivity present in the slices in each superfusion chamber at the end of the experiment varied among the different experiments between 7×10^4 and 10^5 d.p.m. for ^{3}H and between 18×10^3 and 22×10^3 d.p.m. for ^{14}C. The K^+-induced release of radioactivity was determined from the difference between the total overflow in the second and third superfusion samples (that is, from $t = 40$–60 min) and the spontaneous efflux estimated from the radioactivity collected in the first and the fourth 10-min sample. The K^+-induced release was calculated as per cent of the total amount of radioactivity present in the tissue at the time of stimulation. The control values (K^+-induced release in absence of peptides) varied among the experiments between 4.2 ± 0.1% amd 6.3 ± 0.1% (ref. 4) of the total tissue content for ^{3}H-dopamine release and between 5.3 ± 0.2% and 7.7 ± 0.4% (ref. 4) for ^{14}C-acetylcholine release. Release in the presence of peptides has been expressed as per cent of control because experiments varied. The spontaneous efflux of radioactivity during the 10-min period preceding K^+-stimulation, in the absence of peptides, amounted to 0.20–0.25% min^{-1} of total tissue radioactivity (^{3}H as well as ^{14}C). Each value is the mean ± s.e.m. of 12 observations obtained in three separate experiments. Statistical significance of the peptide effects (compared with controls): * $P < 0.001$; † $P < 0.05$; (two-tailed Student's *t*-test).

Initial experiments (Table 1) revealed that dynorphin(1–17) as well as dynorphin(1–13), at a concentration of 0.1 μM inhibited both the spontaneous efflux of tritium (not shown in Table 1) and the K^+-induced release of ^{3}H-dopamine from striatal slices by about 30%, whereas the release of ^{14}C-acetylcholine was not significantly affected. Raising the concentration of dynorphin(1–13) to 1.0 μM did not further increase the inhibition of ^{3}H-dopamine release, indicating that it was already maximal, but at this concentration the peptide also affected K^+-induced ^{14}C-acetylcholine release, reducing it by about 20%.

© 1984 Nature Publishing Group

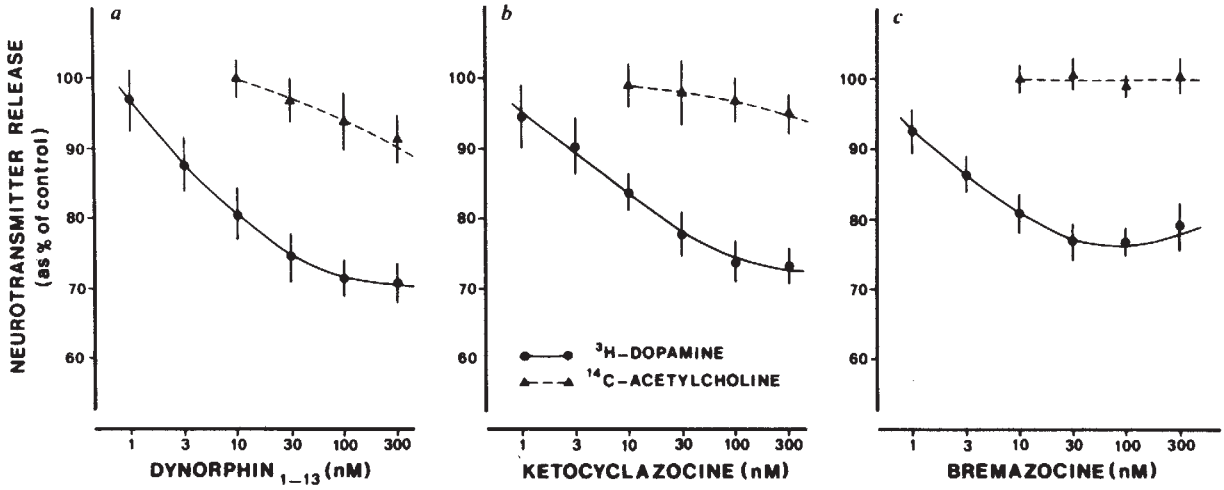


Fig. 1 Effects of the κ-opiate receptor agonists dynorphin(1–13), ketocyclazocine and bremazocine on the K^+-induced release of ^{3}H-dopamine and ^{14}C-acetylcholine, respectively, from rat striatal slices. Preparation and incubation of the slices were carried out as described in the legend to Table 1. The superfusion procedure was analogous to the cumulative dose–response technique developed previously for studying presynaptic receptors in brain slices[33]. After 40 min superfusion with medium containing 20 μM bacitracin the slices were exposed continuously to 15 mM K^+. From $t = 50$ min increasing concentrations of the drug were added cumulatively to the medium with 10-min intervals. The inhibitory effect of the drug was determined by comparing the K^+-induced release of ^{3}H-dopamine in its presence with the release found in control chambers containing slices not exposed to the drug. (See ref. 33 for a detailed description of the method and of the calculation of the data.) Each value is the mean ±s.e.m. of 12–16 observations obtained in 4–5 separate experiments.

In further experiments the concentration–response relationships of the release-inhibiting effects of dynorphin(1–13) and other κ-receptor agonists were determined precisely. Figure 1*a* shows that dynorphin(1–13) inhibits K^+-induced ^{3}H-dopamine release with an EC_{50} (that is, the concentration causing half-maximal inhibition) of approximately 5 nM. An almost identical concentration–response relationship was found for the inhibitory effect of dynorphin on the spontaneous efflux of tritium (not shown). Similarly, the benzomorphans ketocyclazocine and bremazocine also inhibited both the spontaneous and the K^+-induced release of ^{3}H-dopamine with EC_{50} values in the nanomolar range (Figs 1*b*, *c*), whereas ^{14}C-acetylcholine release was not significantly affected up to and including the maximal concentration tested (0.3 μM).

In contrast to dynorphin, [Leu^5]enkephalin and DADLE, at concentrations up to 0.3 μM, did not significantly affect ^{3}H-dopamine release, but these peptides inhibited the K^+-induced release of ^{14}C-acetylcholine with EC_{50} values of approximately 0.1 and 0.03 μM, respectively (Fig. 2). The near-maximal inhibitory effect appeared to be reached at 1.0 μM. At this concentration both peptides caused a slight inhibition (by 10–15%) of the spontaneous efflux of ^{14}C and also affected ^{3}H-dopamine release, reducing it by about 15% (Fig. 2). Note that our finding that [Leu^5]enkaphalin as well as DADLE did not appreciably modulate or, if anything, inhibited (at concentrations ≥ 1 μM) the release of ^{3}H-dopamine contrasts with recent data reported by Lubetzki *et al.*[34]. They found that a number of peptides, including DADLE, caused an increase of the spontaneous release of newly synthesized radiolabelled dopamine from striatal slices and suggested that this effect was mediated by δ-opioid receptors. At present we have no satisfactory explanation for the apparent discrepancy between our finding and theirs.

Morphine (0.1–3.0 μM), the prototype μ-opioid receptor agonist, did not affect the release of ^{3}H-dopamine nor that of ^{14}C-acetylcholine from striatal slices (data not shown). This agrees with findings reported by others[34–38], although at high concentrations (> 10 μM) morphine may cause inhibition of dopamine or acetylcholine release from striatal slices[38,39].

Together our data lead to the tentative conclusion that dynorphin inhibits dopamine release by activating κ-opioid receptors, whereas the inhibitory effect of [Leu^5]enkephalin on acetylcholine release in the corpus striatum is mediated by δ-opioid receptors. The inhibitory effects of dynorphin and [Leu]^{5}enkephalin were relatively insensitive to the opiate antagonist naloxone, which agrees with the view that both κ- and δ-receptors require higher concentrations of naloxone for antagonism than μ-receptors[10,13]. Thus, the inhibition of ^{3}H-dopamine release caused by 0.03 μM dynorphin remained unaffected by 0.01 μM naloxone and was only partly antagonized by 0.1 μM naloxone (Table 2). The inhibition of ^{14}C-acetylcholine release by 0.3 μM [Leu^5]enkephalin too was not affected by 0.01 μM naloxone, but was largely reversed by 0.1 μM of the antagonist. On the other hand, the inhibitory effect of 0.3 μM [Leu^5]enkephalin on the K^+-induced release of ^{3}H-noradrenaline from hippocampal slices, which is mediated by μ-opioid receptors[40,41], appeared to be antagonized largely by only 0.01 μM naloxone (Table 2).

Although the functional implications of these findings remain to be elucidated, to our knowledge this is the first demonstration,

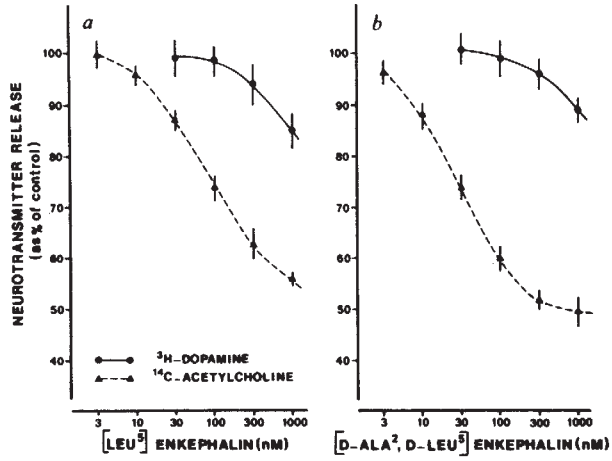


Fig. 2 Effects of the δ-opiate receptor agonists [Leu^5]enkephalin and DADLE on the K^+-induced release of ^{3}H-dopamine and ^{14}C-acetylcholine, respectively, from rat striatal slices. Preparation and incubation of the slices were carried out as described in the legend to Table 1. The superfusion procedure was analogous to the cumulative dose–response technique[33] described briefly in the legend to Fig. 1. Each value is the mean ±s.e.m. of 12–16 observations obtained in 4–5 separate experiments.

© 1984 Nature Publishing Group

Table 2 Antagonism by naloxone of the release-inhibiting effects of dynorphin(1–13) and [Leu^5]enkephalin

In presence of	Release (as % of controls) of: ^{3}H-dopamine (striatum)	^{14}C-acetylcholine (striatum)	^{3}H-noradrenaline (hippocampus)
Dynorphin(1–13) 0.03 μM	70.7±2.7	—	—
Dynorphin(1–13) 0.03 μM+naloxone 0.01 μM	73.6±3.9	—	—
Dynorphin(1–13) 0.03 μM+naloxone 0.1 μM	85.1±4.5†	—	—
[Leu^5]enkephalin 0.3 μM	—	65.0±3.0	62.2±2.5
[Leu^5]enkephalin 0.3 μM+naloxone 0.01 μM	—	69.9±4.3	85.9±3.6*
[Leu^5]enkephalin 0.3 μM+naloxone 0.1 μM	—	89.4±3.8*	93.6±3.2*

The release experiments were carried out as described in the legend to Table 1, except that the striatal slices were labelled with either ^{3}H-dopamine or ^{14}C-choline. Furthermore, hippocampal slices were labelled by incubation with ^{3}H-noradrenaline (specific activity 39 Ci mmol^{-1}; 5 μCi 5 ml^{-1}) and during superfusion release was induced by exposing the slices to 10 mM K^+ for 2 min. The control values of K^+-induced ^{3}H-noradrenaline release (in excess of spontaneous efflux) varied between 3.3±0.1 and 3.9±0.2% (4) of the total tissue content. Naloxone, when used, was added to the superfusion medium 30 min before K^+ stimulation and was present throughout the experiment. At the concentrations used naloxone by itself did not significantly affect the release of ^{3}H-dopamine, ^{14}C-acetylcholine or ^{3}H-noradrenaline. Each value is the mean ±s.e.m. of 12 observations obtained in three separate experiments. Statistical significance of antagonism by naloxone (compared with the peptides alone): $*P<0.01$; $†P<0.05$.

that members of two different opioid peptide families may have differential presynaptic modulatory effects on the release of neurotransmitters in a particular region of the central nervous system. Furthermore, the differential inhibition of striatal ^{3}H-dopamine and ^{14}C-acetylcholine release caused by activation of κ- and δ-receptors, respectively, may be utilized as a functional paradigm in the search for new drugs, particularly antagonists, with selectivity for either of these opioid receptors in the brain.

We acknowledge the generous gifts of dynorphin (Organon), ketocyclazocine (Stirling Winthrop) and bremazocine (Sandoz). We also thank Drs Fred Tilders and Istvan Vermes for their stimulating interest in these studies, Henk Dalecky for typing and Henk Nordsiek for artwork.

Received 19 May; accepted 30 December 1983.

1. Bloom, F. E. *et al. Adv. biochem. Psychopharmac.* **18**, 89–109 (1978).
2. Snyder, S. H. & Childers, S. R. *A. Rev. Neurosci.* **2**, 35–64 (1979).
3. Watson, S. J. *et al. Science* **216**, 85–87 (1982).
4. Vincent, S. R., Hökfelt, T., Christensson, I. & Terenius, L. *Neurosci. Lett.* **33**, 185–190 (1982).
5. Khachaturian, H. *et al. Peptides* **3**, 941–954 (1982).
6. Watson, S. J. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 891–894 (1983).
7. Weber, E. & Barchas, J. D. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1125–1129 (1983).
8. Cuello, A. C. *Br. med. Bull.* **39**, 11–16 (1983).
9. Iwamoto, E. T. & Way, E. L. *Adv. biochem. Psychopharmac.* **20**, 357–407 (1979).
10. Vaughan, P. F. T., *Cell. molec. Biol.* **28**, 369–382 (1982).
11. Langer, S. Z. *Pharmac. Rev.* **32**, 337–362 (1980).
12. Starke, K. *A. Rev. Pharmac. Toxicol.* **21**, 7–30 (1981).
13. Martin, W. R., Eades, C. G., Thompson, J. A., Huppler, R. E. & Gilbert, P. E. *J. Pharmac. exp. Ther.* **197**, 517–532 (1976).
14. Lord, J. A. H., Waterfield, A. A., Hughes, J. & Kosterlitz, H. W. *Nature* **267**, 495–499 (1977).
15. Wüster, M., Schultz, R. & Herz, A. *Biochem. Pharmac.* **30**, 1883–1887 (1981).
16. Wood, P. L. *Neuropharmacology* **21**, 487–497 (1982).
17. Paterson, S. J., Robson, L. E. & Kosterlitz, H. W. *Br. med. Bull.* **39**, 31–36 (1983).
18. Chavkin, C., James, I. F. & Goldstein, A. *Science* **215**, 413–415 (1982).
19. Yoshimura, K., Huidobro-Toro, J. P. & Way, E. L. *Eur. J. Pharmac.* **84**, 17–24 (1982).
20. Dray, A. *Neuroscience* **4**, 1407–1439 (1979).
21. Hattori, T., Singh, U. K. & McGeer, P. L. *Brain Res.* **102**, 164–173 (1976).
22. Bartholini, G., Stadler, H., Gadea Civia, M. & Lloyd, K. G. *Neuropharmacology* **15**, 515–519 (1976).
23. Garattini, S., Pujol, J. F. & Samanin, R. S. (eds) *Interactions Between Putative Neurotransmitters in the Brain* 3–38 (Raven, New York, 1978).
24. Stoof, J. C., Horn, A. S. & Mulder, A. H. *Brain Res.* **196**, 276–281 (1980).
25. Hertting, G., Zumstein, A., Jackish, R., Hoffmann, I. & Starke, K. *Naunyn-Schmiedebergs Archs Pharmak.* **315**, 111–117 (1980).
26. Hughes, J., Kosterlitz, H. W. & Smith, T. W. *Br. J. Pharmac.* **61**, 639–647 (1977).
27. Yang, H. Y., Hong, J. S. & Costa, E. *Neuropharmacology* **16**, 303–307 (1977).
28. Simantov, R., Kuhar, M. J., Uhl, G. R. & Snyder, S. H. *Proc. natn. Acad. Sci. U.S.A.* **74**, 2167–2171 (1977).
29. Goldstein, A. & Ghazarossian, V. E. *Proc. natn. Acad. Sci. U.S.A.* **77**, 6207–6210 (1980).
30. Höllt, V., Haarmann, I., Bovermann, K., Jericz, M. & Herz, A. *Neurosci. Lett.* **18**, 149–153 (1980).
31. Schoffelmeer, A. N. M., Wemer, J. & Mulder, A. H. *Neurochem. Int.* **3**, 129–136 (1981).
32. Mulder, A. H. *Prog. Brain Res.* **55**, 135–156 (1982).
33. Frankhuyzen, A. L. & Mulder, A. H. *Eur. J. Pharmac.* **78**, 91–97 (1982).
34. Lubetzki, C., Chesselet, M. F. & Glowinski, J. *J. Pharmac. exp. Ther.* **222**, 435–440 (1982).
35. Szerb, J. C. *Eur. J. Pharmac.* **29**, 192–194 (1974).
36. Arbilla, S. & Langer, S. Z. *Nature* **271**, 559–561 (1978).
37. Starr, M. *J. pharm. Pharmac.* **30**, 359–363 (1978).
38. Beani, L., Bianchi, C. & Siniscalchi, A. *Br. J. Pharmac.* **76**, 393–401 (1982).
39. Celsen, B. & Kuschinsky, K. *Naunyn-Schmiedebergs Archs Pharmak.* **284**, 159–165 (1974).
40. Taube, H. D., Starke, K. & Borowski, E. *Naunyn-Schmiedebergs Archs Pharmak.* **299**, 123–141 (1977).
41. Göthert, M., Pohl, I. M. & Wehking, E. *Naunyn-Schmiedebergs Archs Pharmak.* **307**, 21–27 (1979).

VIP and noradrenaline act synergistically to increase cyclic AMP in cerebral cortex

Pierre J. Magistretti* & Michel Schorderet*†

* Département de Pharmacologie, Centre Médical Universitaire, 11211 Geneva, Switzerland
† Ecole de Pharmacie, Université de Lausanne, 1000 Lausanne, Switzerland

There is growing evidence that two, or possibly more, neurotransmitters can coexist within the same neurone[1,2]. In particular, the presence of a peptide and a biogenic amine has been demonstrated in the same terminals of central and peripheral neurones[1,2]. These findings have led to the hypothesis that neurotransmitters, coexisting within the same neurones, can interact at pre- or postsynaptic sites in a functionally coordinated manner[1–3]. However, interactions between neurotransmitters contained in distinct neuronal systems terminating within the same region of the central nervous system (CNS) can be envisaged. We have examined this last possibility in the cerebral cortex, an area of the CNS where the two neurotransmitters vasoactive intestinal polypeptide and noradrenaline are contained in separate neuronal systems and where they both stimulate the formation of cyclic AMP[4]. We report here that vasoactive intestinal polypeptide and noradrenaline act synergistically to stimulate the formation of cyclic AMP and that this synergistic interaction is antagonized by the specific α-adrenergic antagonist phentolamine.

Vasoactive intestinal polypeptide (VIP) is a 28 amino acid polypeptide originally isolated from porcine duodenum by Said and Mutt[5] (see ref. 6 for review). It is also present in the CNS, where it has been detected in the highest concentrations in cerebral cortex[7]. VIP immunoreactivity is localized throughout the cerebral cortex to bipolar, radially oriented (that is, in a plane perpendicular to the pial surface), intracortical neurones[7], whose arborization pattern defines partially overlapping, transcortical volumes of approximately 30 μm in diameter[8]. Cellular actions of VIP in the cerebral cortex include the stimulation of cyclic AMP formation[9] and the activation of glycogenolysis[10].

Noradrenaline (NA), another neurotransmitter present in high concentrations in the cerebral cortex[11], is contained in neurones whose morphological characteristics contrast with those of the VIP intracortical system. Thus, the noradrenergic innervation of the cerebral cortex is composed of fine axons, organized predominantly in a plane parallel to the pial surface,

© 1984 Nature Publishing Group

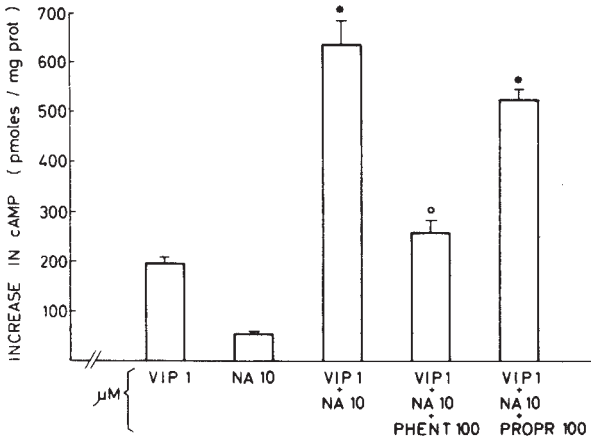


Fig. 1 Synergism between VIP and NA in the stimulation of cyclic AMP formation by mouse cerebral cortical slices. Mice were decapitated, the brain rapidly removed, and the cerebral cortex dissected on ice and placed in a modified Krebs–Ringer bicarbonate buffer (KRG) containing 120 mM NaCl, 5 mM KCl, 2.6 mM $CaCl_2$, 0.67 mM $MgSO_4$, 1.2 mM KH_2PO_4, 3 mM glucose, 27.5 mM $NaHCO_3$ and previously gassed with $O_2:CO_2$ (95:5) to maintain *p*H 7.4. Cortical slices, 250×250 μm, were then prepared on a McIlwain tissue chopper and resuspended on ice-cold KRG (6 ml per cortex). After replacing the buffer, the slices were incubated at 37 °C under vigorous shaking and continuous gassing with $O_2:CO_2$ (95:5). After 15 min the medium was again replaced and aliquots (~2 mg protein per ml) were pipetted into individual tubes. Tubes were gassed with a stream of $O_2:CO_2$ (95:5), capped and incubated at 37 °C. After 30 min, drugs or KRG were added, and the tubes regassed, recapped and incubated for additional 10 min. Antagonists were added 2 min before NA or VIP. The incubation was terminated by a rapid centrifugation followed, after replacement of the supernatant with fresh medium, by a sonication for 5 s. Samples were boiled for 10 min, centrifuged, and cyclic AMP measured in the supernatants as described below. Aliquots of the supernatants (10–100 μl) were lyophilized and resuspended in 20 μl of a Tris/EDTA (0.05 M/4 mM) buffer. Cyclic AMP was then determined according to the saturation assay method described by Brown *et al.*[17]. Experimental values were calculated by computer from a standard curve ranging over 0–16 pmol cyclic AMP. Proteins were determined in the pellets as described by Lowry *et al.*[18]. In all conditions where drugs were tested, cyclic AMP levels were significantly increased over basal levels ($P<0.001$). Absolute cyclic AMP levels were (pmol per mg protein ±s.e.m., $n=8$–17): basal, 13.8±0.91; VIP 1 μM, 211.9±16.7; NA 10 μM, 67.9±4; VIP 1 μM+NA 10 μM, 648.7±54.5; VIP 1 μM+NA 10 μM+phentolamine 100 μM, 274.6±24.4; VIP 1 μM+NA 10 μM+propranolol 100 μM, 538±20. Neither phentolamine nor (±)propranolol antagonized the accumulation of cyclic AMP induced by VIP (not shown). The effects on cyclic AMP levels of VIP plus NA in the presence of the adrenergic antagonists were always compared with those of VIP and NA, alone or combined, within the same experiment. *, Significantly different from the sum of VIP and NA tested separately ($P<0.001$) by Student's *t*-test. 0 = o, Not significantly different from the sum of VIP and NA tested separately by Student's *t*-test.

spanning a vast expanse of cortex, with their cell bodies located in the nucleus locus coeruleus in the brain stem[11–13]. In the cerebral cortex, NA, like VIP, stimulates cyclic AMP formation[14] and promotes glycogenolysis[15] via a calcium-dependent activation of phosphorylase *b* (ref. 16). VIP and NA are thus contained within cortical neuronal systems of a contrasting although complementary nature, and share certain cellular actions[4].

We first examined the effect of a combination of 1 μM VIP and 10 μM NA on the formation of cyclic AMP by mouse cerebral cortical slices. Swiss male albino mice (18–20 g) were used throughout the study and cerebral cortical slices were prepared as previously described[10,15] (see Fig. 1 legend for details). As shown in Fig. 1, the increase in cyclic AMP levels induced by 1 μM VIP in the presence of 10 μM NA is approximately 2.5 times greater than the sum of cyclic AMP levels formed in the presence of the same concentrations of each compound tested separately, thus demonstrating a synergistic effect of VIP and NA in stimulating cyclic AMP formation. Note here that this synergistic interaction is also fully expressed in the presence of the adenosine antagonist theophylline (not shown).

Table 1 Effect of adrenergic agonists on the VIP-induced accumulation of cyclic AMP

Drugs added	Concentration (μM)	Cyclic AMP levels (pmol per mg protein)
None		12.1±0.39
VIP	1	221.9±13.2
Isoprenaline	50	19.0±1.8*
Phenylephrine	50	13.8±0.4
VIP+	1	
isoprenaline	50	222.9±22†
VIP+	1	
phenylephrine	50	412.7±34‡

Methods are described in Fig. 1 legend. Results ±s.e.m. are from 9–19 determinations.

* Significantly different from cyclic AMP levels in the absence of added drugs ($P<0.001$), by Student's *t*-test.

† Not significantly different from 1 μM VIP, by Student's *t*-test.

‡ Significantly different from 1 μM VIP (($P<0.001$), by Student's *t*-test.

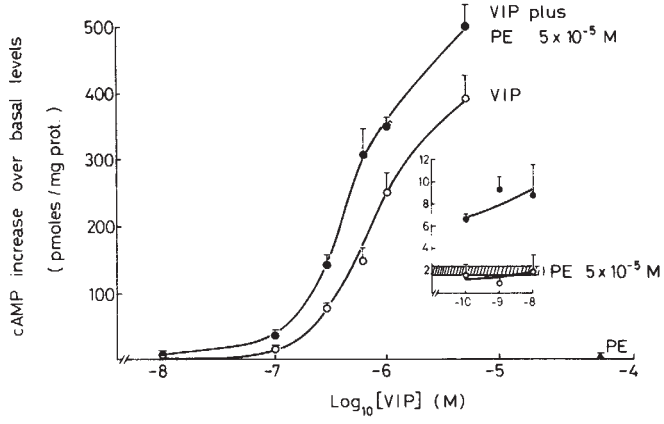


Fig. 2 Effect of 5×10^{-5} M phenylephrine (PE) on the concentration–response curve of the stimulatory effects of VIP on cyclic AMP formation in mouse cerebral cortical slices. Methods are described in Fig. 1 legend. Coordinates for the inset are the same as for the main graph. ●, VIP+5×10^{-5} M PE; ○, VIP alone. The hatched line refers to cyclic AMP levels ±s.e.m. in the presence of 5×10^{-5} M PE. Each point on the curve is the mean ±s.e.m. of four determinations. Basal cyclic AMP levels were 12.06±0.29 pmol per mg protein. The potentiating effect of 5×10^{-5} M PE was highly significant at each VIP concentration (*P* between <0.001 and 0.05).

As shown in Fig. 1, the synergistic effect of 1 μM VIP and 10 μM NA was antagonized by the specific α-adrenergic agonist phentolamine (100 μM) but not by (±)propranolol (100 μM), a specific β-adrenergic antagonist.

Furthermore, as shown in Table 1, the specific α-adrenergic antagonist phenylephrine (PE), at 50 μM, significantly potentiated the stimulating effects of 1 μM VIP on cyclic AMP formation; in contrast, in the presence of the specific β-adrenergic agonist (−)isoprenaline at 50 μM, the levels of cyclic AMP generated by 1 μM VIP were not significantly different from those observed in the presence of 1 μM VIP alone.

Finally, the effects of a fixed concentration (50 μM) of phenylephrine on the concentration–response curve of the stimulatory effects of VIP on cyclic AMP formation were examined. Figure 2 shows, there was a leftward shift of the concentration–response curve, with a significant ($P<0.05$) increase in the maximal effect from 391.9±37.2 to 498.4±36.7 pmol cyclic AMP per mg

© 1984 Nature Publishing Group

protein ±s.e.m. Only a modest decrease in the EC_{50} from 800 nM to 600 nM (assessed by Eadie–Hofstee transformation) was thus observed.

In the present series of experiments we have demonstrated that NA and VIP display a synergistic effect on the formation of cyclic AMP by mouse cerebral cortical slices. Experiments with the adrenergic antagonists phentolamine and (±)- propranolol suggest that activation of α-adrenergic receptors may be involved in this synergism.

In view of the neuronal organization of the cortical VIP- and NA-containing systems, several remarks regarding the cellular actions of the two neurotransmitters can be made. Thus, on the basis of the demonstration of a glycogenolytic action of VIP and NA in the cerebral cortex[10,15], it was proposed that the two neurotransmitter systems could regulate cortical energy metabolism in a complementary manner[4,10]—VIP locally, within radially oriented, circumscribed transcortical volumes, and NA globally, throughout functionally distinct regions of the cerebral cortex. The functional synergism between VIP and NA in stimulating cyclic AMP formation demonstrated in the present report indicates that VIP and NA could in fact act physiologically in concert within a volume of cerebral cortex delineated by the intersection of active noradrenergic fibres and a group of VIP intracortical neurones. This synergism would lead to the generation of a 'metabolic hot spot' within the cortex where cyclic AMP levels would be drastically increased over those in surrounding regions, and activate those intracellular mechanisms that are regulated by cyclic-AMP dependent phosphorylations.

In conclusion, the recently advanced concept that VIP and NA may constitute an example of complementarity between neuronal systems controlling cortical homeostasis[4] can now be extended to one of convergence, at least for those cellular events that are activated by increases in intracellular cyclic AMP. Furthermore, the demonstration of a synergistic effect of VIP and NA on an intracellular event strongly suggests that the two neurotransmitters may act, at least in part, on identical target cells in the cerebral cortex.

We thank Dr Floyd E. Bloom for helpful comments and encouragement, Ms Gisèle Gilliéron for technical assistance, Mr Fred Pillonel for graphical work and Ms Sylvianne Bonnet and Nicole Collet for typing the manuscript. This work was supported in part by SNSF grant 3.345.0.81 to M.S.

Received 3 July; accepted 17 December 1983.

1. Hökfelt, T. *et al. Proc. R. Soc.* **B210,** 63–77 (1980).
2. Hökfelt, T., Johansson, O., Ljungdahl, A., Lundberg, J. M. & Schultzberg, M. *Nature* **284,** 515–521 (1980).
3. Lundberg, J. M. *et al. Neuroscience* **4,** 1539–1559 (1979).
4. Morrison, J. H. & Magistretti, P. J. *Trends Neurosci.* **6,** 146–151 (1983).
5. Said, S. I. & Mutt, V. *Science* **169,** 1217–1218 (1970).
6. Said, S. I. (ed.) *Advances in Peptide Hormones, Research Series* (Raven, New York, 1982).
7. Loren, I. *et al. Neuroscience* **4,** 1953–1976 (1979).
8. Morrison, J. H., Magistretti, P. J., Benoît, R. & Bloom, F. E. *Brain Res.* **292,** 269–282 (1984).
9. Quick, M., Iversen, L. L. & Bloom, S. R. *Biochem. Pharmac.* **27,** 2209–2213 (1978).
10. Magistretti, P. J., Morrison, J. H., Shoemaker, W. J., Sapin, V. & Bloom, F. E. *Proc. natn. Acad. Sci. U.S.A.* **78,** 6535–6539 (1981).
11. Moore, R. Y. & Bloom, F. E. *A. Rev. Neurosci.* **2,** 113–168 (1979).
12. Morrison, J. H., Molliver, M. E., Grzanna, R. & Coyle, J. T. *Neuroscience* **6,** 139–158 (1981).
13. Foote, S. L., Bloom, F. E. & Aston-Jones, G. *Physiol. Rev.* **63,** 844–914 (1983).
14. Nathanson, J. A. *Physiol. Rev.* **57,** 157–256 (1977).
15. Quach, T. T., Rose, C. & Schwartz, J. C. *J. Neurochem.* **30,** 1335–1341 (1978).
16. Ververken, D., Van Veldhoven, P., Proost, C., Carton, H. & De Wulf, H. *J. Neurochem.* **38,** 1286–1295 (1982).
17. Brown, B. L., Albano, J. D. M., Ekins, R. P. & Sgherzi, A. M. *Biochem. J.* **121,** 561–562 (1971).
18. Lowry, O. H., Rosebrough, N. J., Farr, A. L. & Randall, R. J. *J. biol. Chem.* **193,** 265–275 (1951).

A venom peptide with a novel presynaptic blocking action

Lynne M. Kerr & Doju Yoshikami

Department of Biology, University of Utah, Salt Lake City, Utah 84112, USA

The venom of the fish-eating marine mollusc, *Conus geographus*, contains several neurotoxic peptides having different targets[1–5]. A novel peptide has recently been isolated from the venom of *C. geographus* by Drs B. M. Olivera and W. R. Gray and colleagues in our department (in preparation). We report here that this peptide, designated ωCgTX (omega *C. geographus* toxin), irreversibly blocks nerve stimulus-evoked release of transmitter at the frog skeletal neuromuscular junction. Experiments indicate that the toxin acts by preventing Ca^{2+} entry into the nerve terminal during the presynaptic action potential. Consistent with this is the demonstration that ωCgTX also irreversibly attenuates the Ca^{2+} component of the action potential in dorsal root ganglion (DRG) neurones from embryonic chick. ωCgTX thus provides a unique and potentially powerful probe for exploring the presynaptic terminal.

Isolated cutaneous pectoralis muscle–nerve preparations from frog (*Rana pipiens*) were examined by conventional recording techniques. Unless indicated otherwise, the preparations were bathed in normal frog Ringer's consisting of (in mM): NaCl 113, KCl 2, $CaCl_2$ 1.8, HEPES 10, adjusted to *p*H 7.2 with NaOH. For Mg^{2+}-Ringer's, $MgCl_2$ was substituted for $CaCl_2$. Neurones in isolated dorsal root ganglia from chick embryos (20 days old) were examined in a solution consisting of (in mM): NaCl 152, KCl 3, $CaCl_2$ 10, $MgCl_2$ 1, glucose 6, tetraethylammonium chloride (TEA) 10, HEPES (as above). All experiments were performed at room temperature.

When muscle preparations were exposed to ~5 μM ωCgTX, endplate potentials (e.p.ps) evoked by nerve stimulation were completely abolished within 15 min. This block persisted even after the preparations were washed extensively (for example, 8 h at room temperature followed by 48 h at 4 °C). Lower concentrations of toxin (for example 50 nM) were also effective but required longer exposures.

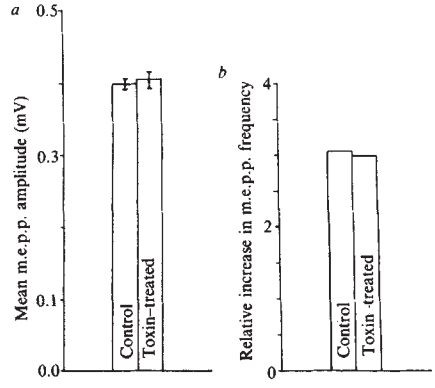


Fig. 1 ωCgTX affects neither the m.e.p.p. amplitudes nor the increase in m.e.p.p. frequency induced by hypertonicity. *a*, Mean amplitudes of m.e.p.ps from a fibre before (left) and after (right) treatment with ωCgTX (error bars denote standard errors; $n = 50$). *b*, Increases in m.e.p.p. frequency induced by shifting from normal Ringer's to a hypertonic (normal Ringer's with 24 mM sucrose) solution before (left) and after (right) treatment with ωCgTX. Higher concentrations of sucrose (for example, 240 mM) increased the m.e.p.p. frequencies of both control and toxin-treated fibres to levels which precluded accurate measurement. The m.e.p.ps of the toxin-treated preparations were assessed only after it was ascertained that the toxin had totally blocked e.p.ps evoked by nerve stimulation.

Although the e.p.ps were totally abolished by ωCgTX, the spontaneous miniature endplate potentials (m.e.p.ps) were unaffected (Fig. 1*a*). Furthermore, muscle action potentials and

© 1984 Nature Publishing Group

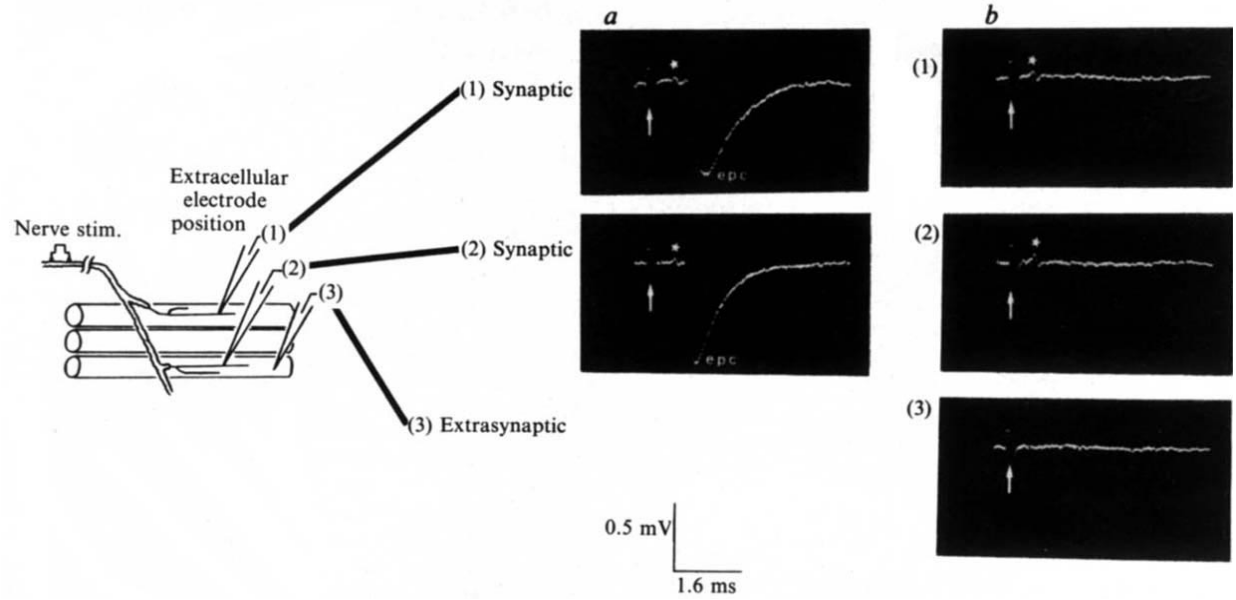


Fig. 2 Conduction of action potentials to, and within, presynaptic nerve terminals is not blocked by ωCgTX. The preparation was in Mg^{2+}-Ringer's, and the tip of an extracellular (1 M $CaCl_2$) microelectrode was placed, under visual guidance (see, for example, ref. 18), on the surface of two fibres either abutting the nerve terminal (synaptic site) or elsewhere (extrasynaptic site) as indicated in the diagram on the left. The traces represent an average of 30 responses to nerve stimulation at 0.5 Hz. *a*, Before toxin treatment, the extracellularly recorded presynaptic action potential (star), and the postsynaptic endplate current (e.p.c.) were recorded from synaptic sites on both fibres (positions 1 and 2) following the stimulus artefact (arrow) (left-hand traces). *b*, After exposure to ωCgTX, no postsynaptic e.p.c. response was recorded although the presynaptic response (star) remained quite evident at both synaptic sites. As a control, the electrode was placed on an extrasynaptic site (position 3) of a fibre, only the stimulus artefact (arrow) was recorded. This indicates that the starred responses in the preceding traces were, indeed, focal responses from the nerve terminal.

contractions in response to direct electrical stimulation were unaffected by the toxin (not shown). Thus, ωCgTX does not have any obvious postsynaptic effects.

ωCgTX does not block release *per se* as is evident from the spontaneous m.e.p.ps which persist after exposure to ωCgTX. Furthermore, transmitter release, in the form of increased m.e.p.p. frequency induced by hypertonicity[6,7], is not blocked by the toxin (Fig. 1*b*). Toxin treatment itself produced no conspicuous changes in m.e.p.p. frequency, but further experiments are necessary to determine whether there might be significant, albeit small, changes.

To test whether ωCgTX acted by blocking the propagation of action potentials to, or within, the presynaptic terminal, we recorded these potentials directly by using focal extracellular electrodes (Fig. 2). The preparation was set up in Mg^{2+}-Ringer's so that synaptic transmission was blocked[8] to prevent the muscle from twitching in response to nerve stimulation. However, the extracellular recording electrode contained 1 M $CaCl_2$, and sufficient Ca^{2+} diffused from it to restore synaptic transmission focally near its tip (see, for example, ref. 9). Before ωCgTX treatment, the extracellular electrode recorded both the impulse in the presynaptic terminal and a postsynaptic endplate current. On exposure to toxin, the presynaptic response persisted whereas, as expected, the postsynaptic response was completely abolished. This indicates that ωCgTX blocks synaptic transmission without interfering with the propagation of action potentials in the nerve terminal.

To gain further insight into the mechanism of action of ωCgTX, we examined its effect before total block of synaptic transmission. After ωCgTX is administered, but before e.p.ps are totally eliminated, there is a period when the responses to nerve stimulation are failures interspersed with e.p.ps whose amplitudes are discrete multiples of m.e.p.p. amplitudes. This condition is similar to that found in normal muscle when the quantal content (m) is low (for example, see ref. 8). The partially blocked preparation can be kept in this condition by simply washing the toxin out. Figure 3*a* shows sample responses from a fibre in such a partially blocked state. Transmitter release in this preparation could be increased by elevating the extracellular

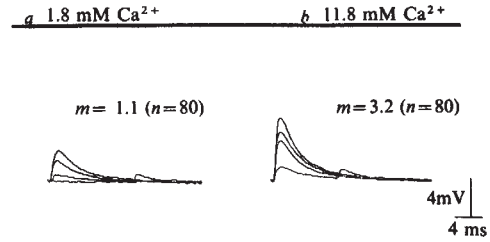


Fig. 3 ωCgTX reduces neurotransmitter release. This preparation was briefly treated with toxin and washed free of unbound toxin when synaptic transmission was only partially blocked. The nerve was stimulated at 1 Hz. *a*, Sample records in normal (1.8 mM) Ca^{2+}. Note the failure and multiquantal responses. The quantal content m was 1.1 (80 trials). *b*, Sample records when [Ca^{2+}] was raised to 11.8 mM, m was increased to 3.2 (80 trials). A spontaneous m.e.p.p. is present on the tail of a response in each frame.

Ca^{2+} concentration ([Ca^{2+}]) above normal (Fig. 3*b*). m, Determined by the ratio of the average e.p.p. to the average m.e.p.p. and m determined by the method of failures (that is, $m = \ln$[no. of trials/no. of failures]) were essentially the same in each condition indicating that release followed Poisson statistics[8]. In this regard, ωCgTX produces a state resembling that normally achieved by reducing the external [Ca^{2+}] (see, for example, ref. 8) or by adding a powerful Ca^{2+} antagonist such as Mn^{2+} (ref. 10).

Taken together, the results indicate that the toxin acts by interfering with Ca^{2+} entry into the nerve terminal during the presynaptic action potential. To substantiate this point, we examined the effect of ωCgTX on the contribution of Ca^{2+} to the action potential in DRG neurones from embryonic chick. In a DRG neurone exposed to TEA, a prolonged plateau is present in the falling phase of the action potential (Fig. 4*a*). Such a characteristic plateau has been shown to be due to Ca^{2+} entry for many neurones, for example, cultured DRG neurones from chick embryos[11]. Ca^{2+} entry apparently is responsible for the plateau in our non-cultured DRG neurones also as the

© 1984 Nature Publishing Group

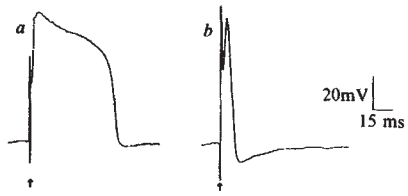


Fig. 4 Effect of ωCgTX on the action potential from a chick DRG neurone. The membrane potential of the cell body was monitored with an intracellular electrode while brief pulses of suprathreshold depolarizing currents were injected (arrow) through the same electrode via a bridge circuit. *a*, Neurone before treatment with ωCgTX—note the prolonged Ca^{2+} plateau following the fast Na^+ spike. *b*, Same neurone after exposure to ωCgTX; the plateau has disappeared although the fast Na^+ spike remains unaltered. The effect of the toxin could not be reversed by washing. The stimulus artefact (arrow) is larger in *b* because a stronger current pulse was used. The preparation was exposed to 10 mM TEA throughout.

plateau is obliterated by the Ca^{2+} antagonist Mn^{2+} (not shown). Significantly, the plateau is also obliterated by treatment with ωCgTX without altering the fast Na^+ spike (Fig. 4*b*). Thus the toxin in some way affects Ca^{2+} entry during an action potential. This could possibly be achieved either by interfering directly with Ca^{2+} channels[12], or by augmenting or hastening the increase of K^+ conductance during the action potential and thereby suppressing the activation of the voltage-sensitive Ca^{2+} channels. Further experiments are needed to distinguish between these possibilities.

The mode of action of ωCgTX is novel and quite distinct from those of previously characterized peptide toxins from bacteria (for example, botulinum toxin) and venoms of snake (for example, β-bungarotoxin) and spider (for example, α-latrotoxin) which act presynaptically (for reviews, see refs 13, 14). In contrast, ωCgTX behaves much like the peptide enkephalin which also blocks the Ca^{2+} component of the action potential in cultured DRG neurones from chick[15] and the release of transmitter at the frog neuromuscular junction[16]. Unlike ωCgTX, however, the effects of enkephalin are readily reversible.

A preliminary survey reveals that ωCgTX also irreversibly blocks synaptic transmission in the spinal cord and sympathetic ganglion from frog[19]. Thus, ωCgTX is effective at several different, if not all, chemical synapses in frog.

We thank Drs B. M. Olivera, W. R. Gray and their associates for providing ωCgTX, and Dr L. M. Okun for critical suggestions and for unpublished data on non-dissociated DRG neurones from embryonic chick. This work was supported by NIH grants NS15543 and NS00465, and a Muscular Dystrophy of America postdoctoral fellowship (to L.M.K.). A preliminary report of this work was presented at the annual meeting of the Society for Neuroscience[17].

Received 24 June; accepted 23 December 1983.

1. Whyte, J. M. & Endean, R. *Toxicology* **1**, 25–31 (1962).
2. Spence, I., Gillessen, D., Gregson, R. P. & Quinn, R. J. *Life Sci.* **21**, 1759–1770 (1977).
3. Gray, W. R., Luque, A., Olivera, B. M., Barrett, J. & Cruz, L. J. *J. biol. Chem.* **256**, 4734–4740 (1981).
4. Clark, C., Olivera, B. M. & Cruz, L. J. *Toxicon* **19**, 691–699 (1981).
5. McManus, O. B., Musick, J. R. & Gonzalez, C. *Neurosci. Lett.* **24**, 57–62 (1981).
6. Fatt, P. & Katz, B. *J. Physiol., Lond.* **117**, 109–128 (1952).
7. Furshpan, E. J. *J. Physiol., Lond.* **134**, 689–697 (1956).
8. del Castillo, J. & Katz, B. *J. Physiol., Lond.* **124**, 560–573 (1954).
9. Katz, B. & Miledi, R. *Proc. R. Soc.* **B161**, 496–503 (1965).
10. Meiri, U. & Rahamimoff, R. *Science* **176**, 308–309 (1971).
11. Dichter, M. C. & Fischbach, G. *J. Physiol., Lond.* **267**, 281–289 (1977).
12. Dunlap, K. & Fischbach, G. *J. Physiol., Lond.* **317**, 519–535 (1981).
13. Kelly, R. B., Deutsch, J. W., Carlson, S. S. & Wagner, J. A. *A. Rev. Neurosci.* **2**, 399–446 (1979).
14. Howard, B. D. & Gundersen, C. B. *A. Rev. Pharmac. Tox.* **20**, 307–336 (1980).
15. Mudge, A. W., Leeman, S. E. & Fischbach, G. D. *Proc. natn. Acad. Sci. U.S.A.* **76**, 526–530 (1979).
16. Bixby, J. L. & Spitzer, N. C. *Nature* **301**, 431–432 (1983).
17. Kerr, L. M. & Yoshikami, D. *Soc. Neurosci. Abstr.* **8**, 496 (1982).
18. Kuffler, S. W. & Yoshikami, D. *J. Physiol., Lond.* **244**, 703–730 (1975).
19. Yoshikami, D., Kerr, L. M. & Elmslie, K. S. *Soc. Neurosci. Abstr.* **9**, 882 (1983).

Inhibition of lung maturation by monoclonal antibodies against fibroblast–pneumonocyte factor

Martin Post, Joanna Floros & Barry T. Smith

Department of Pediatrics, Harvard Medical School, Boston, Massachusetts 02115, USA

Fetal lung maturation, especially the onset of surfactant formation by alveolar type II cells, seems to be regulated by endogenous fetal glucocorticoids[1–3]. Recent studies[4,5] suggest that glucocorticoids do not act directly on the type II cell but rather on the lung mesenchyme. In response to glucocorticoids, the mesenchyme produces and secretes a polypeptide, fibroblast–pneumonocyte factor, which in turn stimulates surfactant synthesis by the alveolar type II cell[6]. We report here the generation of hybridomas secreting monoclonal antibodies to rat lung fibroblast–pneumonocyte factor. Two monoclonal antibodies studied in detail reduced the cortisol-stimulated synthesis of saturated phosphatidylcholine in organotypic cultures of fetal rat lung cells and blocked the stimulatory effect of fibroblast-pneumonocyte factor in type II cells from these cultures. When embryonic chicks were injected on day 15 of incubation with either monoclonal antibody, they showed on days 20 and 21 biochemical evidence of delayed lung maturation as compared with controls. These effects were organospecific. Our observations support a physiological role for fibroblast–pneumonocyte factor in prenatal lung maturation.

We generated hybridomas that produced antibodies to fibroblast–pneumonocyte factor (FPF) by fusion of spleen cells from immunized mice with NSl myeloma cells. Partially purified FPF was prepared by gel filtration of conditioned medium (FCM) from cortisol-treated fetal lung fibroblasts on Biogel P-60 (ref. 5). The factor (200 μg of protein) was further purified by SDS-polyacrylamide gel electrophoresis[5] and duplicate gels were then either stained for protein with Coomassie brilliant blue G or fixed in 1 M acetic acid. The fixed gel was sliced into 10 5-mm pieces and each piece was homogenized in 1.5 ml saline by pushing the gel–saline mixture through sequentially smaller needles (18–26 gauge).

Ten female BALB/c mice were immunized with the corresponding homogenates. Each mouse received, in 4-week intervals, three intraperitoneal injections (0.5 ml) of homogenate from one of the bands. Analysis at 12 weeks demonstrated that sera from the two mice injected with gel fractions containing polypeptides of apparent molecular weight 5,000–10,000 blocked the effect of FCM in the bioassay[7]. In contrast, preimmune sera from these mice did not block the FCM effect. These mice received a booster injection of 0.5 ml FCM intraperitoneally. Three days later the mice were killed and their spleens removed for hybridization.

The spleen cells were fused with NS1 myeloma cells using polyethylene glycol as described by Kohler and Milstein[8] to produce hybridomas. Supernatant fluids from the resulting hybridoma cells were tested for the presence of antibody in a bioassay, as previously described[7]. From two independent fusions we obtained hybridomas in 83 wells. Of 22 hybrid cells which secreted antibodies against FPF, 2 hybridomas, termed H1 and 2D2, which produced antibodies with the highest biological activity as assessed by the bioassay, were successfully grown intraperitoneally in BALB/c mice. Both hybridoma antibodies were found to be IgM as determined by double immunodiffusion using specific antisera. We purified the antibodies from ascites fluid collected from inoculated BALB/c mice by ammonium sulphate precipitation and Sephadex G-200 chromatography. The IgM-containing fraction was dialysed against 5 mM Tris-HCl *p*H 7.5, which caused precipitation of IgM. The precipitate was redissolved in 0.1 M Tris-HCl *p*H 8.6, and dialysed against Tris-buffered saline *p*H 8.6.

© 1984 Nature Publishing Group

Table 1 Effect of anti-FPF antibodies on saturated phosphatidylcholine (SPC) formation by fetal rat lung type II cells in organotypic cultures

Medium addition	Choline incorporated (d.p.m. per 10^6 cells, mean ± s.d., $n = 4$)
None	4,318 ± 110
Cortisol	7,871 ± 1,113*
Cortisol + antibody 2D2	3,973 ± 587
Cortisol + antibody H1	3,938 ± 933

Organotypic cultures were prepared with lung cells from fetal rats at 19 days gestation and maintained as described previously[9]. The medium was replaced with minimal essential medium containing 1 μCi ml^{-1} ^{3}H-choline (final specific activity 0.14 Ci $mmol^{-1}$), the control condition, and either 10^{-7} M cortisol, 10^{-7} M cortisol plus antibody 2D2 (final concentration 43 μg ml^{-1}) or 10^{-7} M cortisol plus antibody H1 (final concentration 38 μg ml^{-1}). The incubations were terminated after 20 h by isolating type II cells from the organotypic cultures[16]. Lipids were extracted and analysed[9]. In separate experiments, the antibodies did not reduce activity in cultures which had not been exposed to cortisol (data not shown).

* $P = 0.0007$.

These monoclonal IgM antibodies blocked the cortisol-stimulated incorporation of labelled choline into saturated phosphatidylcholine (SPC) by alveolar type II cells in organotypic cultures from fetal rat lung (Table 1). This observation indicates that FPF is involved in the effect of cortisol on surfactant synthesis by fetal lung type II cells in this heterogeneous system.

Furthermore, when alveolar type II cells were isolated from organotypic cultures and grown as monolayers[9], one of the antibodies (2D2) blocked the FPF effect (Table 2). In addition, this blocking activity could be removed by prior incubation with rabbit anti-mouse IgM and its specificity is demonstrated by the observation that IgM purified from non-immune mouse serum has no effect on FPF action in the bioassay (Table 2).

Because IgM antibody does not cross the mammalian placenta, and as maternal laparotomy with direct fetal injection itself accelerates fetal lung maturation, we selected embryonic chicks as a model for further investigating the role of FPF in lung maturation *in vivo* in baseline conditions. Recent studies[10,11] have demonstrated that avian lung maturation is under hormonal control and that a variety of fetal (embryonic) lung fibroblasts (including chick) produce apparently identical FPF-like activity[12].

Monoclonal antibody (2D2) was injected intraperitoneally into chicks on day 15 of incubation (hatching 22 days). A dose was chosen, based on the *in vitro* observations and assuming even distribution throughout the egg contents, which would give 50% inhibition of type II cell response to FPF. On day 21 of incubation, the chicks were killed and lung maturity was studied.

Table 2 Effect of anti-FPF antibody (2D2) on the stimulation of SPC synthesis by alveolar type II cells in response to cortisol–fibroblast conditioned medium

Medium	Addition	Choline incorporated (d.p.m. per 10^6 cells, mean ± s.d., $n = 4$)
MEM	None	2,100 ± 135
MEM	Anti-FPF IgM (1 μg ml^{-1})	2,218 ± 74
FCM	None	4,490 ± 344*
FCM	Anti-FPF IgM (1 μg ml^{-1})	2,390 ± 331
FCM	Anti-FPF IgM (1 μg ml^{-1}) + rabbit anti-mouse IgM (25 μg ml^{-1})	4,359 ± 533*
FCM	Mouse IgM (1 μg ml^{-1})	4,196 ± 409*

Type II cells were isolated from organotypic cultures, grown as monolayers[9], and studied as indicated in Table 1. FCM, cortisol–fibroblast conditioned medium; MEM, minimal essential medium.

* $P < 0.001$.

Table 3 Effect of monoclonal antibody (2D2) on lung maturation *in ovo*

	Individual phospholipids expressed as % of total phospholipids (mean ± s.d., $n = 3$)				
	PC	PE	Sph	PI/PS	PG
Control	45.3 ± 2.5	22.0 ± 7.0	24.7 ± 4.5	7.0 ± 0.7	3.0 ± 1.4
Experimental	42.1 ± 6.2	23.1 ± 8.7	25.4 ± 1.9	7.6 ± 4.5	3.1 ± 0.7
	SPC as % of phosphatidylcholine				
Control	46.5 ± 3.6				
Experimental	32.9 ± 6.7*				

White Leghorn eggs were maintained from day 1 at 38 °C. On day 15 the chicks were injected intraperitoneally with either antibody (278 μg protein in 0.1 ml Tris-buffered saline, *p*H 8.6) or vehicle (0.1 ml). On day 21 chick lungs were removed, weighed and homogenized in cold methanol. Lipids were extracted from lung tissue and separated into individual classes by TLC. SPC was isolated by the osmium tetroxide method and TLC. Phospholipids were measured by assaying phospholipid phosphorus[9]. Abbreviations: PC, phosphatidylcholine; SPC, saturated phosphatidylcholine; PE, phosphatidylethanolamine; Sph, sphingomyelin; PI, phosphatidylinositol; PS, phosphatidylserine; PG, phosphatidylglycerol.

* $P < 0.05$.

No significant differences ($P > 0.05$, Student's *t*-test) were observed (experimental versus control, means ± s.d.) in lung weight (10.0 ± 1.4 versus 9.0 ± 1.4 mg per g body weight) or total lung phospholipids (8.17 ± 0.19 versus 7.92 ± 1.96 μmol per g lung weight). The pulmonary phospholipid profile of control lungs was comparable with that of previous studies[13], indicating that the injection did not affect lung maturation (Table 3). We observed a significantly lower content of saturated phosphatidylcholine, the major surface-active component of pulmonary surfactant, in lungs from chicks injected with antibody 2D2 (Table 3). The per cent saturation of phosphatidylcholine found in the experimental lung on day 21 of incubation was similar to that reported previously[13] for lung from chicks on day 14. A separate experiment showed that, compared with controls, labelled choline incorporation into SPC of chick lung was decreased in chicks injected on day 15 of incubation with antibody H1 (Table 4).

The organ specificity of the antibodies was determined by measuring enzyme activities in two other foregut derivatives, the duodenum and liver. Preliminary evidence suggests that the effect of glucocorticoids on epithelial differentiation in both organs involves mesenchymal–epithelial interactions[14,15]. No significant differences were observed in duodenal alkaline phosphatase (67.1 ± 36.8 versus 38.5 ± 22.3 μmol per min per mg protein, experimental versus control) and liver tyrosine

Table 4 Effect of monoclonal antibody H1 on SPC synthesis in embryonic chick lung

	Choline incorporation (d.p.m. per mg protein, mean ± s.d., $n = 3$)		
	PC	SPC	SPC as % of PC
Control	9,755 ± 1,576	3,005 ± 424	30.9 ± 2.5
Experimental	9,056 ± 646 •	1,815 ± 197*	19.7 ± 2.9†

Chicks were injected intraperitoneally on day 15 of incubation with either antibody (415 μg protein in 0.1 ml Tris-buffered saline, *p*H 8.6) or vehicle (0.1 ml). On day 20 lungs were removed, sliced and incubated for 2 h in medium 199 containing 1.5 μCi of ^{3}H-choline (final specific activity 0.14 Ci $mmol^{-1}$) at 38 °C. The incubations were terminated by placing the incubation vessels on ice and washing the tissues with 10 ml of cold medium. Lipids were extracted and separated by TLC and SPC was isolated[9]. The radioactivity incorporated into PC and SPC was determined.

* $P < 0.01$; † $P < 0.005$.

© 1984 Nature Publishing Group

aminotransferase (7.58 ± 2.41 versus 6.56 ± 2.25 nmol per min per mg protein).

We conclude that the present observations strongly endorse the view that glucocorticoid-stimulated lung maturation is mediated by mesenchymal–epithelial interactions, in the form of fibroblast–pneumonocyte factor. Furthermore, they demonstrate that, in addition to the ability of exogenous FPF to accelerate lung maturation[5], this factor is involved in the normal course of development of the lung. The organ specificity of this peptide[4,12] is supported by these observations.

We thank Dr A. Abbas for advice and assistance, Dr A. Schwartz for suggesting the immunization technique, Mr D. Righi for performing the tyrosine aminotransferase assays, and Dr H. Feldman for statistical advice. A. Boyer, A. Barsoumian and S. Smith provided technical assistance. This work was supported by NIH grants HL27372, HL25907 and HD14534; and by the Carl W. Walter Fellowship Fund.

Received 23 August; accepted 17 December 1983.

1. Smith, B. T. & Bogues, W. G. *Pharmac. Ther.* **9**, 51–74 (1980).
2. Kitterman, J. A. *et al. J. appl. Physiol.* **51**, 384–390 (1981).
3. Ballard, P. L. in *Lung Development: Biological and Clinical Perspectives* (ed. Farrell, P. M.) 205–253 (Academic, New York, 1982).
4. Smith, B. T. in *Neonatal Intensive Care* (ed. Stern, L.) 25–32 (Masson, New York, 1978).
5. Smith, B. T. *Science* **204**, 1094–1095 (1979).
6. Smith, B. T. in *The Biology of Normal Human Growth* (eds Ritzen, M. *et al.*) 157–162 (Raven, New York, 1981).
7. Smith, B. T. & Sabry, K. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1951–1954 (1983).
8. Kohler, G. & Milstein, C. *Nature* **256**, 495–497 (1975).
9. Post, M., Torday, J. S. & Smith, B. T. *Expl Lung Res.* (in the press).
10. Marin, L., Tordet, C. & Dameron, F. L. *Anat. Embryol.* **152**, 223–234 (1978).
11. Torday, J. S., Zinman, H. M. & Nielsen, H. C. *Biol. Neonate* (in the press).
12. Smith, B. T. in *The Physiological and Biochemical Basis for Perinatal Medicine* (eds Monset-Couchard, M. & Minkowski, A.) 54–58 (Karger, Basel, 1981).
13. Hylka, V. W. & Doneen, B. A. *J. exp. Zool.* **220**, 71–80 (1982).
14. Smith, B. T. & Zinman, H. M. *Pediat. Res.* **16**, 117 (1982).
15. Dow, K. E., Sabry, K. & Smith, B. T. *Cell Tissue Res.* **231**, 83–91 (1983).
16. Sevanian, A., Kaplan, S. A. & Barrett, C. T. *Biochim. biophys. Acta* **664**, 498–512 (1981).

Chromosome translocation can occur on either side of the c-*myc* oncogene in Burkitt lymphoma cells

Mary Davis*, Susan Malcolm* & Terence H. Rabbitts†

* MRC Human Genetic Diseases Research Group, Biochemistry Department, Queen Elizabeth College, Campden Hill Road, London W8 7AH, UK

† MRC Laboratory of Molecular Biology, University Medical School, Hills Road, Cambridge CB2 2QH, UK

In Burkitt lymphoma cells reciprocal chromosomal translocations are observed between the long arm of chromosome 8 (8q24) and either the long arm of chromosome 14 (14q32), the short arm of chromosome 2 (2p12) or the long arm of chromosome 22 (22q11)[1,2]. Gene mapping studies have shown that c-*myc*, the cellular homologue of the viral *myc* oncogene, is localized at 8q24 (ref 3–5) and that the three immunoglobulin gene loci map at the breakpoints involved in the translocation: immunoglobulin heavy-chain genes are located at 14q32 (refs 6, 7), κ light chains at 2p12 (ref. 8) and λ light chains at, or close to, 22q11 (refs 9, 10). This correlation suggests an association of immunoglobulin gene rearrangements with the occurrence of these specific translocations. By using *in situ* hybridization we have examined the translocation point with respect to c-*myc* in two cell lines containing 2;8 translocation, and report here that the c-*myc* gene remains on the chromosome 8 involved in the reciprocal exchange with chromosome 2. We have also confirmed that the c-*myc* gene moves from the translocated chromosome 8 in a cell line having a 8;14 translocation. These results show that chromosomal breakage can occur on either side of the c-*myc* gene in Burkitt lymphoma cells.

A close association of the c-*myc* gene with immunoglobulin heavy-chain gene sequences after translocation has been found in several cases[5,11–14]. Analysis of somatic cell hybrids formed from fusions of Burkitt lymphoma lines P3HR1 and Daudi with mouse cells showed that the c-*myc* gene was included in the piece of chromosome 8 translocated to chromosome 14 ($14q^+$) while V_H genes (but neither C_μ nor C_γ) were reciprocally translocated from chromosome 14 to 8 ($8q^-$)[15]. As no data exist regarding the fate of the c-*myc* gene in the variant Burkitt lymphoma lines carrying 2;8 translocations, we have done *in situ* hybridization studies on two cell lines, JI and LY91 (ref. 2), which carry 2;8 translocations. The probe used for these experiments (M13JI7myc1.5) has been described previously[13] and contains a 1.5-kilobase (kb) *Sac*I fragment encompassing exon 2 from a genomic human c-*myc* gene. Chromosome preparations made from Burkitt lymphoma lines JI and LY91 (2;8) were prebanded with Lipsol and photographed. After denaturation the chromosomes were hybridized with the nick-translated c-*myc* probe. After autoradiography, the same cells were located and photographed[16]; this allowed all chromosomes to be identified and removed any possible bias in the selection of cells. Other chromosome abnormalities apart from the 2;8 translocation were seen in many of the cells examined. In JI the most common variants were a marker chromosome 6 and trisomy 11.

Figure 1 shows histograms of grains counted over the relevant chromsomes after *in situ* hybridization to both 2;8 translocation lines. The results show an accumulation of grains only over the distal portion of the long arm of both pairs of chromosomes 8, and an approximately equal number of grains on the normal and abnormal chromosomes ($8q^+$) 8 in both cell lines. The short arm of chromosome 2, the recipient of the material translocated from chromosome 8 in the reciprocal translocation (chromosome $2p^-$), is not labelled. Only one grain on an abnormal chromosome 2, was found in all the karyotypes analysed. In the experiment using JI cells, 30% of all grains observed were found over 8q and in LY91 the fraction was 13%; 82% of all cells contained a labelled chromosome 8 in JI hybridizations and 64% in the case of LY91. These results show conclusively that, in both cell lines studied, the breakpoint in chromosome 8 is distal to the c-*myc* gene, and that this gene is not translocated to chromosome 2. Figure 2 shows the results of a similar experiment using the same hybridization probe to detect c-*myc* genes in the Daudi cell line, which carries an 8;14 translocation. Although the normal chromosome 8 was labelled significantly, no grains were found on the abnormal chromosome 8 ($8q^-$) and instead grains were found on the abnormal chromosome 14 ($14q^+$). This result confirms previous studies[15] which showed that in cells having 8;14 transolcations the c-*myc* gene is translocated to chromosome 14 in the reciprocal exchange.

The results presented here demonstrate that the chromosomal breakpoint in Burkitt lymphoma marker chromosomes can occur on either side of the c-*myc* gene on chromosome 8. The general features of these rearrangements are illustrated in Fig. 3. As it has been demonstrated by somatic cell genetics[15] and *in situ* hybridization (Fig. 2) that since the c-*myc* gene moves onto the abnormal chromosome $14q^+$, the new association of immunoglobulin and c-*myc* genes on $14q^+$ in a head-to-head arrangement[11,13] establishes the transcription orientation of both c-*myc* and immunoglobulin C_H genes on their respective chromosomes—the C_H genes are oriented towards the telomere of chromosome 14 and the c-*myc* gene towards the centromere of chromosome 8 (Fig. 3*b*). The *in situ* hybridization experiments shown in Fig. 1 demonstrate that, at least in the two cell line having 2;8 translocations used here, the c-*myc* gene does not move from the $8q^+$ abnormal chromosome. As the 3′ end of the c-*myc* gene faces the telomere of the normal chromosome 8, the results show that the breakpoint in these cell lines is beyond the 3′ side of c-*myc* (see Fig. 3*a*).

Clearly, there is extreme variability in the position of the chromosomal translocation point in Burkitt lymphoma cells as it appears to occur both upstream and downstream of the c-*myc* gene. The precise point of breakage is also very variable, with breakage within c-*myc* itself being rare. Further, the downstream translocation points in JI and LY91 chromosomes do

© 1984 Nature Publishing Group

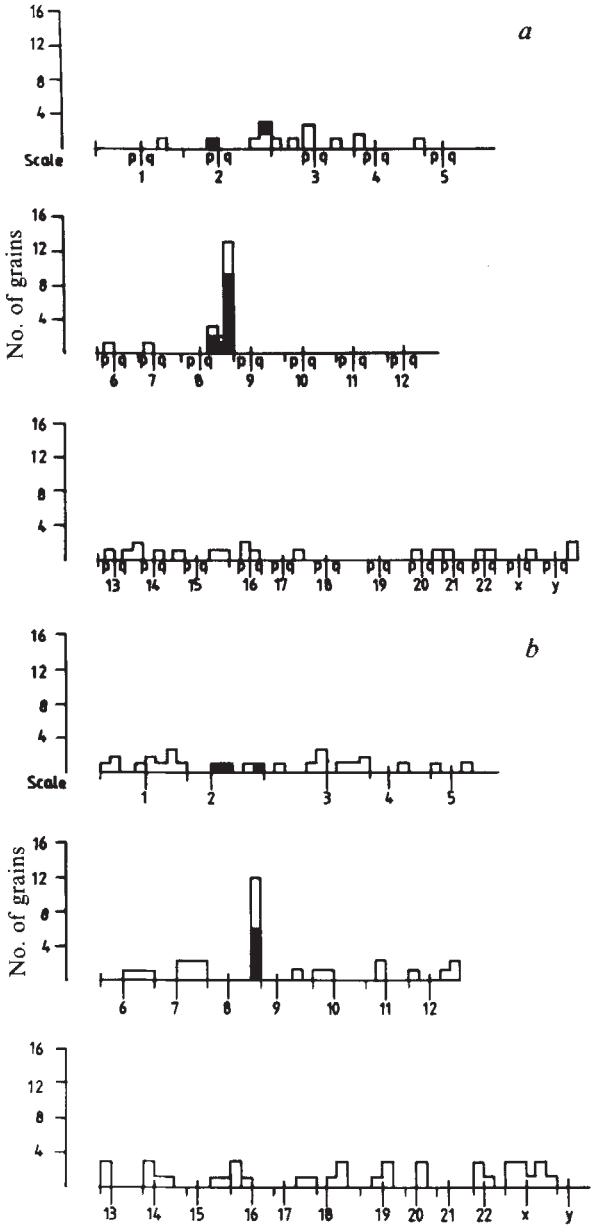


Fig. 1 *In situ* hybridization of c-*myc* genes in Burkitt lymphoma cell lines carrying 2;8 translocations. Distribution of grains over metaphase chromosomes from JI (*a*) or LY91 (*b*) cells. The data are shown as histograms in which the open boxes represent the grains over normal chromosomes; the solid boxes denote grains over the abnormal chromosomes and above these are represented the grains over the normal allelic chromosome, superimposed on the abnormal count. Eleven cells were analysed in each case; all grains were counted over each chromosome and the grain counts were corrected for chromosome length. The DNA hybridization probe was labelled with ^{3}H-dCTP to a specific activity of 1.3×10^8 d.p.m. μg^{-1} by nick-translation. Chromosomes were denatured in 60% formamide, 0.1 mM EDTA, 5 mM HEPES pH 7.0 at 55 °C for 7 min, then 20 ng of boiled probe in 20 μl of hybridization buffer (50% formamide, 0.6 M NaCl, 5 mM HEPES, 1 mM EDTA, 10% dextran sulphate pH 7.6) were applied to each slide and hybridized at 43 °C for 16 h. Slides were washed in 2×SSC at room temperature overnight, then autoradiographed using Ilford K2 nuclear emulsion. Exposure time, 17 days.

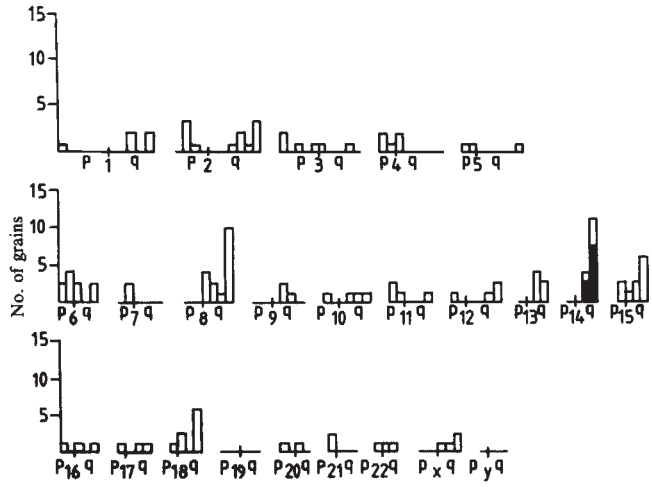


Fig. 2 *In situ* detection of c-*myc* genes in Daudi cells (t8;14). The solid columns indicate grains counted over abnormal chromosomes; see Fig. 1 legend. These are superimposed on the normal chromosome grain count for clarity. No grains were observed on any of the abnormal chromosomes 8. Slides were exposed for 13 days. Eleven slides were analysed and, as in Fig. 1, the grain counts were corrected for chromosome length.

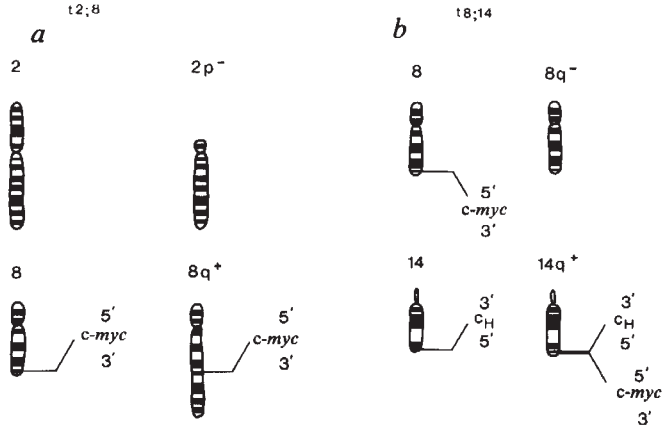


Fig. 3 Diagrammatic representation of translocation in cells carrying t2;8 and t8;14. *a*, Chromosomes 2 and 8, showing the positions of the c-*myc* gene. *b*, Chromosomes 8 and 14, showing the positions and orientations of c-*myc* and immunoglobulin C_H genes.

not affect restriction fragments observed with c-*myc* probes in filter hybridizations (T.H.R., unpublished), thus the breaks must occur at least 6 kb downstream of c-*myc*. It has been shown recently that the c-*myc* gene similarly remains on chromosome 8 in 8;22 Burkitt's lymphoma[17], so both types of variant Burkitt lymphoma exhibit this phenomenon. Several possible consequences of the translocation for the c-*myc* gene have been proposed (reviewed in refs 18, 19).

The intriguing situation in Burkitt's lymphoma is the extreme variability of the chromosomal breakpoint. Although it has been argued that translocations involving c-*myc* result in enhanced levels of transcription of this gene[20], we and others find no such significant enhanced transcriptional activity of the translocated c-*myc* gene compared with lymphoblastoid cell lines[21], nor does an immunoglobulin transcriptional enhancer seem to be generally involved in c-*myc* gene transcription[22]. Furthermore, in view of the finding that the translocation point in Burkitt's lymphoma can occur upstream or downstream of the c-*myc* gene, a common mechanism of oncogenesis involving only elevated levels of c-*myc* mRNA seems unlikely. It has been reported, however, that the translocated c-*myc* allele is preferentially transcribed[23] so that the oncogenic activation of the c-*myc* gene may be the maintenance, rather than elevation, of transcription of this gene. We have recently found that the untranslated exon 1 of the translocated c-*myc* gene in Burkitt's lymphoma often undergoes nucleotide changes, which may explain the loss of transcriptional control resulting from translocation possibly by prevention of repressor protein binding to exon 1 (T.H.R., P. H. Hamlyn and R. Baer, in preparation). Finally, it has been

© 1984 Nature Publishing Group

shown recently that somatic mutation can affect the c-*myc* gene in chickens after avian leukaemia virus insertion[24] and in man after translocation[25]. This mutation may serve to amplify the oncogenic effect of c-*myc* gene translocation by altering the activity of the protein product of this gene.

This work was funded in part by a Leukaemia Research Fund grant to S.M. We thank G. Lenoir for providing the cell lines JI and LY91, and we acknowledge the technical assistance of A. Forster.

Received 16 November 1983; accepted 17 January 1984.

1. Manolov, G. & Manolova, Y. *Nature* **237**, 33–34 (1974).
2. Bernheim, A., Berger, R. & Lenoir, G. *Cancer Genet. Cytogenet.* **3**, 307–315 (1981).
3. Dalla-Favera, R. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 7824–7827 (1982).
4. Neel, B. G., Jhanwar, S. C., Chaganti, R. S. K. & Hayward, W. S. *Proc. natn. Acad. Sci. U.S.A.* **79**, 7842–7846 (1982).
5. Taub, R., *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 7837–7841 (1982).
6. Croce, C. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **76**, 3416–3419 (1979).
7. Hobart, M. J. *et al. Ann. hum. Genet.* **45**, 331–335 (1981).
8. Malcolm, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 4957–4961 (1982).
9. Erikson, J., Martinis, J. & Croce, C. M. *Nature* **294**, 173–175 (1981).
10. de la Chappelle, A. *et al. Nucleic Acids Res.* **11**, 1133–1144 (1983).
11. Dalla-Favera, R., Martinotti, S., Gallo, R. C., Erikson, J. & Croce, C. M. *Science* **219**, 963–967 (1983).
12. Adams, J. M., Gerondakis, S., Webb, E., Corcoran, L. M. & Cory, S. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1982–1986 (1983).
13. Hamlyn, P. H. & Rabbitts, T. H. *Nature* **304**, 135–139 (1983).
14. Battey, J. *et al. Cell* **34**, 779–787 (1983).
15. Erikson, J., Finan, J., Nowell, P. C. & Croce, C. M. *Proc. natn. Acad. Sci. U.S.A.* **79**, 5611–5615 (1982).
16. Davis, M., Malcolm, S., Hall, A. & Marshall, C. J. *EMBO J.* **2**, 2281–2283 (1983).
17. Croce, C. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 6922–6926 (1983).
18. Rabbitts, T. H. *Molec. Biol. Med* **1**, 275–281 (1983).
19. Leder, P. *et al. Science* **222**, 765–771 (1983).
20. Erikson, J., Ar-Rushdi, A., Drivinga, H. L., Nowell, P. C. & Croce, C. M. *Proc. natn. Acad. Sci. U.S.A.* **80**, 820–824 (1983).
21. Maguire, R. T., Robins, T. S., Thorgeirsson, S. S. & Heilman, C. A. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1947–1950 (1983).
22. Rabbitts, T. H., Forster, A., Baer, R. & Hamlyn, P. H. *Nature* **306**, 806–809 (1983).
23. Ar-Rushdi, A. *et al. Science* **222**, 390–393 (1983).
24. Westaway, D., Payne, G. & Varmus, H. E. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
25. Rabbitts, T. H., Hamlyn, P. H. & Baer, R. *Nature* **306**, 760–765 (1983).

Chromosome localization in normal human cells and neuroblastomas of a gene related to c-*myc*

Manfred Schwab*, Harold E. Varmus*, J. Michael Bishop*, Karl-Heinz Grzeschik†, Susan L. Naylor‡, Alan Y. Sakaguchi‡, Garrett Brodeur§ & Jeffrey Trent||

* Department of Microbiology and Immunology and George W. Hooper Foundation, University of California School of Medicine, San Francisco, California 94143, USA
† Institut für Humangenetik, Westfälische Wilhelms-Universität, Münster, FRG
‡ Department of Human Genetics, Roswell Park Memorial Institute, New York State Department of Health, Buffalo, New York 14263, USA
§ Departments of Pediatrics and Genetics, Washington University School of Medicine, St Louis, Missouri 63110, USA
|| Department of Internal Medicine, University of Arizona College of Medicine, Tucson, Arizona 85724, USA

Cellular oncogenes comprise a class of genes whose aberrant expression or function may be involved in the development of tumours[1]. Indeed, several naturally occurring animal and human tumours are associated with consistent alterations in the structure or genomic position of particular cellular oncogenes[2–4]. Recently, we isolated a DNA segment having limited similarity to c-*myc* (termed N-*myc*) from a human neuroblastoma cell line[5]. Although N-*myc* was present as a single copy in normal cells, it was selectively amplified up to 140-fold in tumour cells from human neuroblastomas[5]. Now, we have used somatic cell hybrids to show that N-*myc* is normally localized on the distal short arm of chromosome 2, and *in situ* hybridization to localize N-*myc* to chromosome 2p23–24. Further, *in situ* hybridization localizes amplified N-*myc* in neuroblastoma cells to homogeneously staining regions (HSRs) on different chromosomes. Thus, our results suggest that amplification and translocation of N-*myc* may be interrelated processes associated with human neuroblastoma, and demonstrate that the site of N-*myc* amplification is quite variable and bears no apparent relationship to either the normal single-copy locus or recognized sites of non-random chromosome alteration in human neuroblastoma.

For chromosomal assignment of N-*myc* we used mouse × human somatic cell hybrids differing with respect to their human chromosome content. The DNA of the hybrid cells was cut with the restriction endonuclease *Eco*RI and analysed by Southern blotting procedures[6]. For detection of N-*myc* we used a ^{32}P-labelled *Eco*RI/*Bam*HI insert of clone pNb-1 previously isolated from neuroblastoma line Kelly[5]. As reported previously, the Nb-1 probe detects a 2.0-kilobase pair (kbp) fragment in human DNA digested with restriction endonuclease *Eco*RI[5] (Fig. 1*a*). In mouse DNA digested with *Eco*RI the probe detects a major 16-kbp and a minor 15-kbp fragment (Fig. 1*b*), reflecting the evolutionary conservation of N-*myc*.

We have surveyed the DNA of a panel of mouse × human somatic cell hybrids for chromosomal assignment of human N-*myc*. Human N-*myc* is coordinately present only in hybrids carrying a karyotypically normal chromosome 2 and chromosome 2 markers: isozyme markers isocitrate dehydrogenase 1 (IDH1; q32 → qter), malate dehydrogenase 1 (MDH1; p23) and acid phosphatase 1 (ACP1; p25) (data not shown). Thus N-*myc* appears to be asyntenic with at least 16 oncogenes previously mapped to other human chromosomes. The sole exception at the moment is c-*fos*, which is localized on chromosome 2q (I. Verma, personal communication).

For a more defined localization of N-*myc* on chromosome 2, we used three additional hybrid lines which contain only parts of chromosome 2. The two hybrids A9 Call 1-9-3 and A9 Call 1-13 were constructed by fusing human Call cells [46 X, Y, t(2; 6) (p23; p23)] with mouse A9 fibroblasts[7]. Hybrid A9 Call 1-9-3 contains the translocation (2; 6) and has retained from chromosome 2 the major region qter → p23 (2qter →

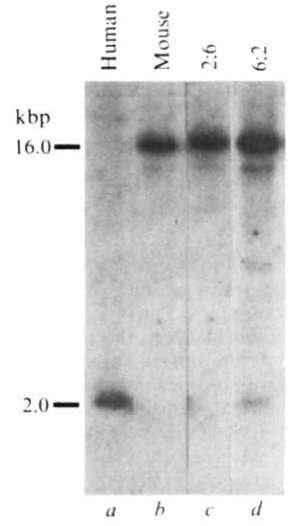


Fig. 1 Identifying N-*myc* in the DNA of mouse × human somatic cell hybrids. *a*, Human skin fibroblasts containing normal chromosome set; *b*, mouse control; *c*, hybrid line (A9 Call 1-9-3) containing a 2;6 translocation (2qter → p23 :: 6p23 → pter; compare with Fig. 2); *d*, hybrid line (A9 Call 1–13) containing a 6;2 translocation. DNA was isolated according to standard procedures[5], cut with restriction endonuclease *Eco*RI and analysed by Southern blotting. As a probe we used the *Eco*RI/*Bam*HI insert of clone pNb-1 previously isolated from neuroblastoma line Kelly[5]. Conditions of hybridization were as described previously[5]; the formamide concentration in the hybridization mixture was 50%. The relatively faint signal obtained in lane *d* for the 2.0-kbp signature fragment is due to the fact that the human 6;2 chromosome is present at a sub-diploid level in the hybrid cells. Lane *c*, the 2.0-kbp fragment was not detectable even in highly overexposed autoradiographs.

© 1984 Nature Publishing Group

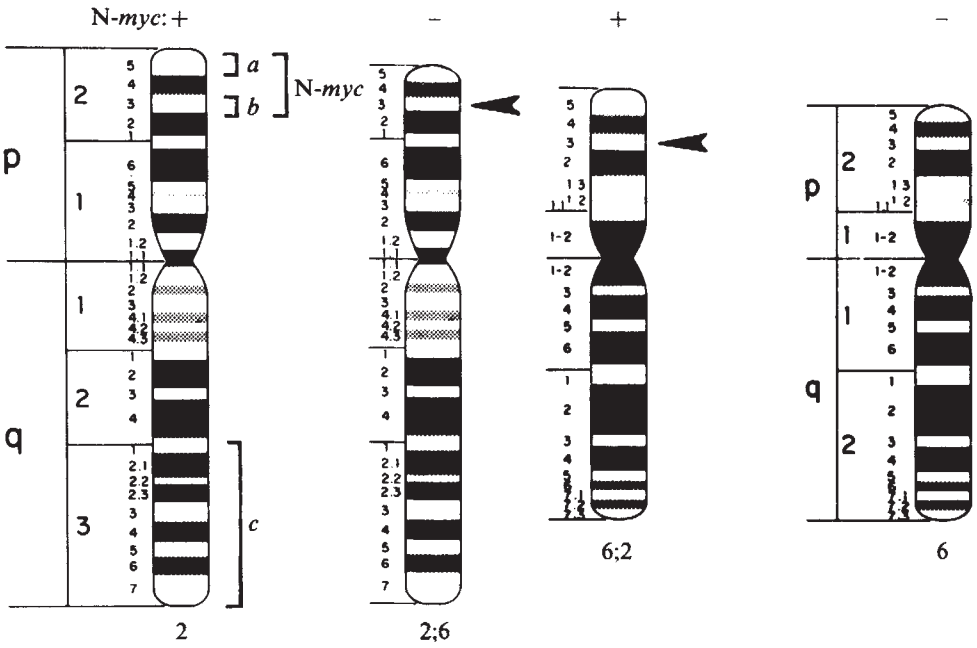


Fig 2 Diagram showing the translocation involving chromosomes 2 and 6 in Call A9 hybrids that were used to further localize N-*myc* on chromosome 2. The 2;6 translocation (2qter→ p23::6p23→ pter) is present in A9 Call 1-9-3. The reciprocal 6;2 translocation (6qter→ p23::2p23→ pter) is present in A9 Call 1–13. Arrowheads indicate the chromosome breakpoints (p23) on chromosomes 2 and 6. Normal chromosomes 2 and 6 are shown for comparison. The location of known chromosome 2 short and long arm markers and of N-*myc* are indicated by brackets: *a*, acid phosphatase (ACP1; p25); *b*, malate dehydrogenase (MDH1;p23); *c*, isocitrate dehydrogenase 1 (IDH1; q32→ qter). Gene symbols and regional assignments of markers are from ref. 26. +/− represent the presence or absence of N-*myc* in the hybrid cell line carrying the particular chromosome, as determined with the Nb-1 probe by Southern blotting of genomic DNA digested with restriction endonuclease *Eco*RI.

p23 :: 6p23→pter; Fig. 2). Hybrid A9 Call 1-13 contains the reciprocal translocation (6; 2) and has retained from chromosome 2 only a small fragment p23→pter (6qter→p23::2p23→ pter). Neither of the two hybrid lines contains an additional normal chromosome 2. DNA from A9 Call 1-9-3, in which all but the distal short arm of chromosome 2 is present, does not show the 2.0-kbp signature fragment (Fig. 1*c*). However, DNA from A9 Call 1-13 [t(6:2)], which has the small fragment p23→ pter of chromosome 2, does show the 2.0-kbp fragment (Fig. 1*d*). Thus N-*myc* is localized to the distal short arm of chromosome 2 (2p23→ pter). An additional hybrid line that we used, WIL8X, contains the chromosome 2 marker ACP1 but lacks the markers MDH1 and IDH1; this is the result of a breakage of chromosome 2 with retention in the hybrid of the segment encoding ACP1 (ref. 8). Hybrid WIL8X contains human N-*myc* (data not shown). This result confirms that N-*myc* is normally localized on the short arm of chromosome 2 in the p23→ pter region.

For independent confirmation and more detailed localization of N-*myc* within the distal short arm of chromosome 2, we used *in situ* hybridization. We examined the autoradiographic images of 100 mitotic cells following *in situ* hybridization with ^{3}H-labelled N-*myc* probe by methods described previously[9]. Significant label was observed on the near-distal short arm of chromosome 2. Of 57 total chromosomal grains localized to chromosome 2, 49.1% were observed on 2p, with 35% (20/57) specifically localized to band p23 or p24. Figure 3*a* gives an example of grain localization to 2p from two different cells. Examination of grain distribution on chromosome 2 from 100 cells analysed indicated significant clustering of grains to chromsome 2p23 or 24 (35%) (Fig. 3*b*). This finding is consistent with our results for mouse × human somatic cell hybrids. It is of interest that an independently derived, as yet uncharacterized, sequence which was shown to be amplified in the neuroblastoma cell line IMR-32, has also recently been demonstrated to localize to the short arm of chromosome 2 (ref. 10).

Table 1 Relationship between N-*myc* DNA detectable by Southern blotting and by *in situ* hybridization

Cell line	Tumour type	Ref.	Fold amplification*	HSR-bearing chromosome	Percentage grain localization
NMB	NB	7	100–120	13p11	68.6
NGP	NB	7	120–140	4p16	57.6
				12q13	4.7
NLF	NB	†	20–25	9q13	22.2
NCI-417	SCLC	26	0‡	1q42	3.0

NB, neuroblastoma; SCLC, small cell lung cancer. * From ref. 5.; † G.B., unpublished data; ‡ J.T., unpublished data.

Three HSR-bearing human neuroblastoma tumour cell lines NMB[11], NGP[11] and NLF (G.B., unpublished data) were also analysed following *in situ* hybridization with ^{3}H-labelled N-*myc* probe. Previous cytogenetic analysis had demonstrated HSRs in these lines on four different chromosomes: NMB, 13p11; NGP, 4p16, and 12q13; NLF, 9q13 (Fig. 3*c*). All the lines carry amplified N-*myc*[5]. Autoradiographic images of 300 metaphases were examined from these lines. Figure 3*c* presents the results from a representative experiment illustrating schematically the grain distribution to HSRs from a total of 62 cells. We found significant hybridization to three of the four HSR regions: NMB, 13p11 (68.6%); NGP, 4p16 (57.6%); and NLF, 9q134 (22.2%). The absolute number of autoradiographic grains over the HSRs was rougly proportional to the level of N-*myc* amplification determined previously by agarose gel electrophoresis and Southern blotting (Table 1). In the neuroblastoma cell line NGP, which contains two HSRs, only the HSR on 4p16 demonstrated significant localization of autoradiographic grains (Fig. 3*c*). The second HSR present on chromsopme 12q13 failed to yield grain counts above background (4.7%). As a control for the specificity of hybridization, ^{3}H-labelled N-*myc* was hybridized *in situ* to the human HSR-bearing small cell carcinoma cell line, NCI-417, which contains amplified copies of the cellular oncogene, c-*myc*[12]. As expected by the relative lack of similarity between c-*myc* and N-*myc* observed by Southern blotting[5], no binding of the ^{3}-labelled N-*myc* probe above background was observed (Table 1).

In this study, we have regionally mapped N-*myc* within the normal karyotype to 2p23 or 24. In contrast to cellular oncogenes specifically related to chromosomal breakpoints in certain cancers[2–4], in human neuroblastomas 2p23 or 24 is rarely involved in either numeric or structural chromosome alterations[13,14]. Deletions or rearrangements involving 1p32 → pter have been detected in ~70% of 23 neuroblastomas and tumour-derived cell lines[15], whereas structural rearrangements of 2p have been observed in only 4% of cases from the same series.

Our results show that HSRs in cells of human neuroblastomas contain amplified copies of N-*myc*. Of the three neuroblastoma

© 1984 Nature Publishing Group

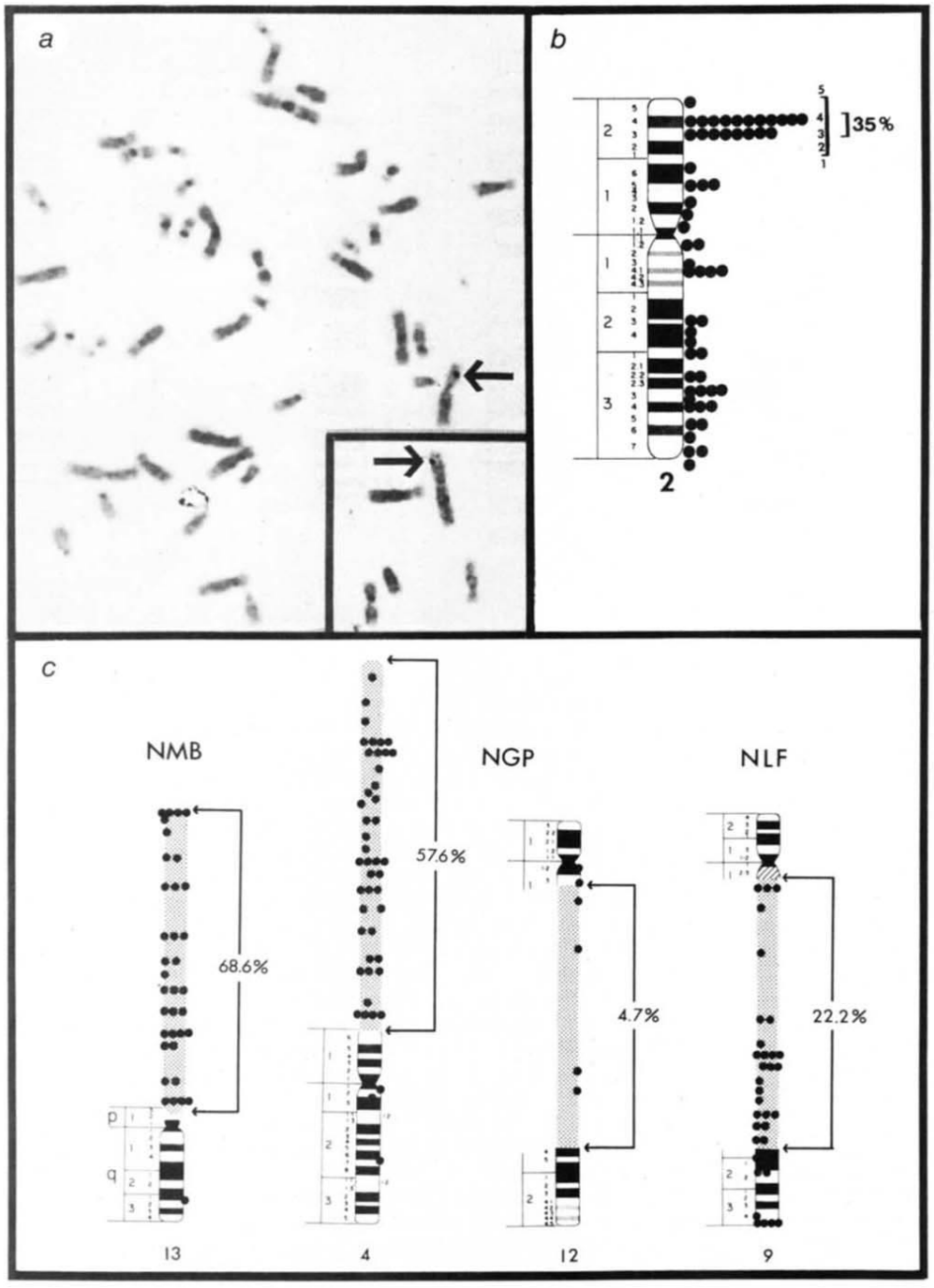


Fig. 3 Cytological evidence for localization of N-*myc* to human chromosome 2p23 or 24 in normal cells, and to three HSR-bearing marker chromosomes in neuroblastoma cell lines. *a*, Representative label (arrows) on the near-distal short arm of chromosome 2 (band 23 or 24) from two different cells hybridized *in situ* with the ^{3}H-labelled N-*myc* probe. Cells were hybridized for 16 h at a DNA concentration of 0.755 mg ml^{-1}, followed by exposure for 28 days and autoradiographic development[9]. *b*, Schematic representation of G-banded mid-metaphase chromosome 2 depicting all grains observed from 100 cells analysed *in situ* with ^{3}H-labelled N-*myc* DNA. Of the grains detected on chromosome 2, 35% were localized to p23 or 24. *c*, Schematic representation of G-banded mid-metaphase chromosomes from three neuroblastoma cell lines depicting all chromosomal grains observed from 62 cells hybridized *in situ* with ^{3}H-labelled N-*myc* DNA. Stippled regions represent the relative size and location of HSRs from the NMB, NGP and NLF cell lines (see Table 1). Total chromosomal grains localized to HSRs were: NMB, 13p11 (68.6%); NGP, 4p16 (57.6%) and 12q13 (4.7%); NLF, 9q13 (22.2%).

lines examined, we found N-*myc* amplification in HSRs on three different chromosomes (4, 9 and 13, Table 1). The failure of these amplifications to correlate with the chromosomal locus of N-*myc* on normal cells (2p23 or 24) is not surprising based on studies of other amplified domains. In methotrexate-resistant human and murine cell lines carrying amplified gene copies for dihydrofolate reductase, the chromosomal site of amplification is highly variable and often seems to 'evolve' into multiple HSR sites following increasing time/drug exposure[16,17]. A similar mechanism may operate in human neuroblastomas where HSRs have been observed on 18 of the 24 different human chromosomes[14]. To date, only one cell line (LA-N-1)[18] analysed contains an HSR on chromosome 2p23. Unexpectedly, the NGP cell line has an additional HSR (at 12p13) which does not appear to hybridize with the N-*myc* probe by *in situ* hybridization. It may be of considerable interest to determine what additional sequences are amplified at this site.

It is unclear whether the chromosomal material is translocated directly. Several neuroblastoma cell lines carry double-minutes[5] DMs and these extrachromosomal elements have also been found in primary neuroblastoma[19–21]. Studies by others on amplification of drug resistance genes[22] and by us on amplification of cellular oncogenes[15,23,24] have shown that both HSRs and DMs are the cytological manifestations of amplified DNA. It seems likely, therefore, that DMs in neuroblastoma cells also contain amplified N-*myc*[5], and it is possible that the translocation of N-*myc* in neuroblastoma cells is mediated by reintegration of DMs at various positions in the cellular genome.

At least three genetic alterations are associated at high consistency with human neuroblastoma: translocation of N-*myc*, amplification of N-*myc*, and previously reported but as yet uncharacterized deletions within the short arm of chromosome1 (ref. 13). In addition, a mutant version of a cellular *ras* gene has been identified in at least one case[25]. Tumorigenesis is generally regarded as a multi-step process. The data now available suggest that this is true for human neuroblastoma, and that several of the steps involved are genetic abnormalities affecting cellular oncogenes.

© 1984 Nature Publishing Group

These studies were supported by grants from the NIH, NCI and the ACS. M.S. is a Heisenberg-fellow of the Deutsche Forschungsgemeinschaft. We thank Bernhard Zabel for helpful discussions.

Note added in proof: Since this manuscript was submitted, Kohl *et al.*[27] also reported amplification of N-*myc* in neuroblastoma cell lines. Together with our previously presented results[5], this brings the number of neuroblastoma cell lines reported positive for N-*myc* amplification to 14 (out of 18 tested), in addition to 5 out of 8 fresh neuroblastoma tumours[5,24]. Kohl *et al.* also localized N-*myc* to the short arm of chromosome 2 and determined in one cell line double minutes as the site of amplification.

Received 20 October 1983; accepted 12 January 1984.

1. Bishop, J. M. *A. Rev. Biochem.* **52**, 801–817 (1982).
2. Klein, G. *Cell* **32**, 311–315 (1983).
3. Rowley, J. D. *Nature* **301**, 290–291 (1983).
4. Yunis, J. Y. *Science* **221**, 227–235 (1983).
5. Schwab, M. *et al. Nature* **305**, 245–248 (1983).
6. Southern, E. M. *J. molec. Biol.* **98**, 503–517 (1975).
7. Grzeschik, K.-H., Mayerova, A., Bender, K. & Johannsmann, R. *Cytogenet. Cell Genet.* **22**, 412–414 (1978).
8. Owerbach, D. *et al. Somat. Cell Genet.* **7**, 359–369 (1981).
9. Trent, J. M., Olson, S. & Lawn, R. M. *Proc. natn. Acad. Sci. U.S.A.* **79**, 7809–7813 (1982).
10. Kanda, N. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 4069–4073 (1983).
11. Brodeur, G. M., Sekhon, G. S. & Goldstein, M. N. *Cancer* **40**, 2256–2263 (1977).
12. Little, C. D., Nau, M. M., Carney, D. N., Gazdar, A. F. & Minna, J. D. *Nature* **306**, 194–196 (1983).
13. Brodeur, G. M. *et al. Cancer Res.* **41**, 4678–4686 (1981).
14. Biedler, J. L., Meyers, M. B. & Spengler, B. A. *Cell. Neurobiol.* **4**, 267–307 (1983).
15. Alitalo, K., Schwab, M., Lin, C. C., Varmus, H. E. & Bishop, J. M. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1707–1711 (1983).
16. Harper, M. A. & Kellems, R. E. *Cancer Bull.* **33**, 43–38? (1981).
17. Trent, J. M. in *Gene Amplification* (ed. Schimke, R. T.) 99–105 (Cold Spring Harbor Laboratory, New York, 1982).
18. Seeger, R. C. *et al. Cancer Res.* **37**, 1364–1371 (1977).
19. Levan, A., Manolov, G. & Clifford, P. *J. natn. Cancer Inst.* **41**, 1377–1387 (1968).
20. Franck, F. & Lampert, F. *Onkologie* **5**, 268–272 (1982).
21. Sandberg, A. A., Sakurai, M. & Holdsworth, R. N. I. *Cancer* **29**, 1671–1679 (1972).
22. Schimke, R. T., Kaufman, R. J., Alt, F. & Kellems, R. F. *Science* **202**, 1051–1055 (1978).
23. Schwab, M., Alitalo, K., Varmus, H. E., Bishop, J. M. & George, D. *Nature* **303**, 497–501 (1983).
24. Schwab, M., Alitalo K., Varmus, H. E. & Bishop, J. M. *Cold Spring Harb. Conf. on Cell Proliferation and Cancer* **11**, (in the press).
25. Shimizu, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**. 2112–2116 (1983).
26. *Cytogenet. Cell Genet.* (in the press).
27. Kohl, N. E. *et al. Cell* **35**, 359–367 (1983).

Tumour induction in the rodent *Mastomys natalensis* by activation of endogenous papilloma virus genomes

Eberhard Amtmann*, Manfred Volm† & Klaus Wayss†

* Institute for Virus Research and † Institute for Experimental Pathology, German Cancer Research Centre, 6900 Heidelberg, FRG

Specific DNA sequences from human papillomavirus have recently been detected in carcinomas from epidermodysplasia verruciformis patients[1,2], and in vulvar[3] and cervical[4,5] carcinomas but the role of papilloma viruses in the aetiology of these tumours is unclear. Indeed, little is known about the mechanisms that convert benign papillomas into malignant tumours and it is not even possible to distinguish between inactive passengers and viruses involved in tumour induction. Here, we describe an animal system that permits an analysis of the interaction of papilloma virus genomes with carcinogenic agents at the molecular level. In our colony of *Mastomys natalensis* (a close relative of the rat family), we have found extrachromosomal papilloma virus genomes persisting in a variety of tissues such as skin, muscle, liver and colon. With the ageing of the animals, the average copy number of viral DNA in skin cells increases and virus-producing tumours begin to appear in *Mastomys* at about 1 year old. This process is drastically enhanced by chronic treatment with a tumour promoter and transcription of the viral genomes has been found to be correlated with tumour formation.

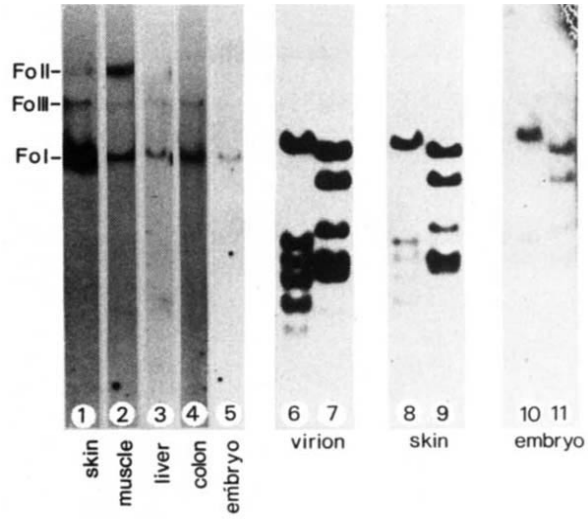


Fig. 1 Detection of MnPV-specific DNA sequences in normal tissues of *Mastomys natalensis*. The DNA preparations in lanes 1–5 were uncleaved; those in lanes 6, 8 and 10 were cleaved with *Hpa*II, and those in lanes 7, 9 and 11 were cleaved with *Hha*I. The DNA samples were isolated from: lane 1, skin; lane 2, muscle; lane 3, liver; lane 4, colon; lane 5, embryo; lanes 6 and 7, mouse 3T3 cells mixed with 200 pg MnPV virion DNA; lanes 8 and 9, skin; lanes 10 and 11, total embryo.
Methods: A 6-month-old animal was killed a few days before the end of the gestation period and samples (0.5 g) of different tissues were frozen in liquid nitrogen. After disruption of the tissue, DNA was purified by phenol and chloroform extractions as described[10]. From each DNA preparation 20 μg were electrophoresed on 1% agarose gels for 16 h at 3 V cm^{-1} and transferred to nitrocellulose filters. MnPV-specific DNA sequences were revealed on X-ray films after hybridization to ^{32}P-labelled MnPV DNA which was cloned in pBR322 at its single *Hin*dIII site. Filters were exposed for 10 days. The details of the procedure were described elsewhere[11].

In a previous report, spontaneously arising skin tumours were described in 3% of the animals of a *M. natalensis* colony[6]. *M. natalensis* papillomavirus (MnPV) was identified as the aetiological agent of these tumours[7,8]. As in our colony of *Mastomys*, animals developed tumours independently from contacts with tumour-bearing animals, we used the Southern blot hybridization procedure to search for latent MnPV genomes. In all DNA preparations from the skin of 5 females and 5 males, we detected extrachromosomal papillomavirus genomes. In four cases, non-skin tissues also contained viral genomes. Figure 1 shows the results obtained from a pregnant animal. In the DNA samples isolated from skin (lane 1), muscle (lane 2), liver (lane 3) and colon (lane 4), superhelical, relaxed circular and some linear MnPV DNA can be seen. Interestingly, embryos that were surgically removed from the uteri of pregnant animals also contained extrachromosomal MnPV genomes (Fig. 1, lane 5). The authenticity of the signals is shown after cleavage with the multi-cut restriction endonucleases *Hpa*II and *Hha*I (Fig. 1, lanes 6–11). Similar results were obtained in several other animals tested. Oesophagus, larynx and bronchial tissues were also frequently found to contain MnPV genomes, while lymphocytes were in no case positive.

The age of the animals was found to be an important factor in tumour formation. Tumours were never observed in animals that were younger than 50 weeks old. In contrast, among 10 animals aged 16 months, 8 were bearing tumours. To follow the kinetics of a putative accumulation of viral DNA during ageing, 50 animals from several litters were homogeneously mixed, and groups of 10 were killed at intervals of 16 weeks. The number of viral genomes per skin cell was determined for each animal by densitometer scanning of the autoradiographs of Southern blots. In young animals (16 weeks old) and in embryos we found 0.1–10 genome equivalents. Only 9 out of

© 1984 Nature Publishing Group

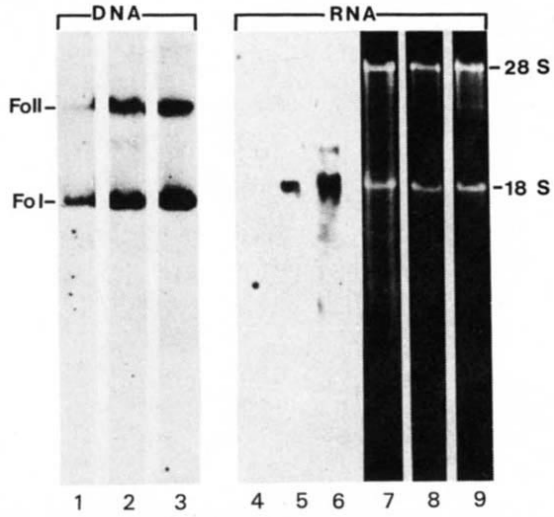


Fig. 2 Detection of MnPV-specific RNA sequences in tumorous and acanthotic tissues. Lane 1, DNA, and lanes 4 and 7, RNA isolated from normal skin of a tumour-free animal; lane 2, DNA, and lanes 5 and 8, RNA isolated from skin of a tumour-free animal (by histological examination, multiple areas of acanthosis and hyperplasia were detected in an adjacent sample); lane 3, DNA, and lanes 6 and 9 RNA isolated from a keratoacanthoma.
Methods: DNA and RNA was simultaneously prepared from skin or tumour samples of 64-week old animals as described elsewhere[10]. MnPV-specific DNA sequences were revealed by Southern blot analysis as described in Fig. 1 legend. The X-ray film was exposed to the filter for 10 min (lanes 1–3). The RNA samples isolated from the same tissues were electrophoresed on a 1.4% agarose gel supplemented with 15 mM methylmercuric hydroxide at 3 V cm^{-1} for 16 h. After electrophoresis the RNA was stained with ethidium bromide and visualized under UV light (lanes 4–6). The RNA was transferred from the gel to a nitrocellulose filter and hybridized to ^{32}P-labelled MnPV DNA as described previously[12]. The X-ray film was exposed for 12 days (lanes 4–6). The positions of the 28S and 18S ribosomal RNAs in the RNA gel are indicated.

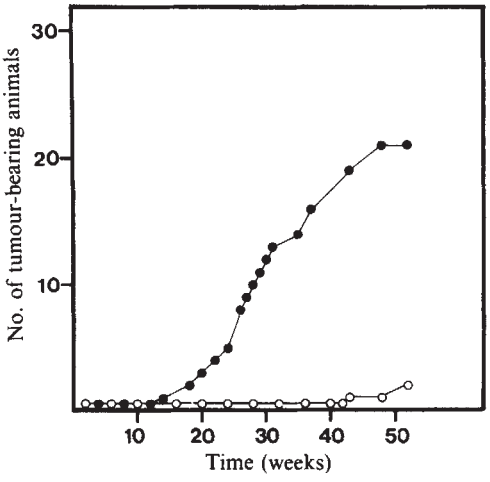


Fig. 3 Tumour induction by tropical treatment with TPA. TPA (40 nmol) was administered to the dorsal skin of 6-week-old *Mastomys* twice weekly (29 animals). After 2 weeks, the dose was reduced to 20 nmol per treatment because of toxic effects. A control group (28 animals) was treated with 0.2 ml acetone. The animals were examined for tumours every week. All tumours arising were identified histologically as keratoacanthomas.

14 embryos and 6 out of 10 skin preparations from the 16-week-old animals contained viral DNA above the level of detection. As the animals aged, the relative amounts of viral DNA increased drastically. The average number of genome equivalents per cell rose from 0.1 at 16 weeks, 10 at 32 weeks, 100 at 48 weeks and 800 at 64 weeks up to 3,000 at 80 weeks. It is important to note that in some cases several thousand copies of viral DNA were detected in skin tissue that was proven to be normal by histological examination. By Northern blot hybridizations, we detected viral transcripts in all of 10 tumours tested (Fig. 2, lane 6) and in preneoplastically altered skin (Fig. 2, lane 5) such as hyperplasia or acanthosis (two samples). MnPV-specific RNA sequences were detected in none of eight normal skin samples (Fig. 2, lane 4), although >1,000 viral genomes per cell were estimated to be present (Fig. 2, lane 1). The integrity of the RNA preparations was monitored by ethidium bromide staining of the gel (Fig. 2, lanes 7–9).

The tumour promoter 12-*O*-tetradecanoylphorbol-13-acetate (TPA) was shown to interact with unexpressed papillomavirus genomes *in vitro*[9]. After 14 weeks of chronic treatment of young *Mastomys* with TPA, one animal developed a tumour at the site of TPA application. After 44 weeks, when the first tumour appeared in the control group, 19 (66%) of the treated animals were already bearing tumours (Fig. 3) that were histologically indistinguishable from spontaneous tumours. As in untreated animals, we detected viral transcripts exclusively in tumour tissue (data not shown).

By Southern blot hybridization of skin DNA from animals that were treated for 8 weeks with TPA, we found a 100-fold stimulation of the viral DNA content in the skin cells. In six control animals (17 weeks old), ~0.2 MnPV genome was present per skin cell, while six TPA-treated animals contained, on average, 50 genomes. When treated and untreated skin portions from the same animal were compared, the viral DNA content was found to be stimulated by TPA by 30–250-fold.

The data presented here show that MnPV-specific DNA sequences persist as plasmids in normal tissues of *M. natalensis*, and that the viral DNA content in skin cells is increased either by ageing or by chronic treatment with TPA. Furthermore, transcription of the MnPV genomes was consistently noted in conjunction with the development of tumours.

By *in situ* hybridization with sections of normal skin from tumour-bearing and tumour-free animals, we could detect no cells containing remarkably high copy numbers of viral DNA, nor could we detect any cells that synthesized virions by indirect immunofluorescence with sera from tumour-bearing animals. The synthesis of virus was exclusively confined to the keratinized layers of tumours.

Although we failed to detect viral transcripts in normal tissue despite the presence of viral DNA, a temporary or low-level expression of the papillomavirus DNA may have escaped detection. In contrast, MnPV-specific RNA sequences were consistently found in hyperplasia, acanthosis and/or tumours. The mechanism responsible for the effects of TPA observed here *in vivo* may be similar to that responsible for the temporary induction of papillomavirus gene expression reported to occur in tissue culture cells treated with TPA[9].

We thank W. Baader and J. Boldrin for technical assistance, H.-J. Hacker for thin sections and G. Sauer for helpful discussions and critical reading of the manuscript.

Received 7 December 1983; accepted 10 January 1984.

1. Ostrow, R. S. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 1634–1638 (1982).
2. Pfister, H., Gassenmeier, A., Nünberger, F. & Stuttgen, G. *Cancer Res.* **43**, 1436–1441 (1983).
3. Zachow, K. R. *et al., Nature* **300**, 771–773 (1982).
4. Gissmann, L. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 560–563 (1983).
5. Green, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 4437–4441 (1982).
6. Rudolf, R. & Thiel, W. *Zentbl. VetMed.* **A23**, 429–441 (1976).
7. Rudolf, R. and Müller, H. *Zentbl. VetMed.* **B23**, 143–150 (1976).
8. Müller, H. & Gissmann, L. *J. gen. Virol.* **41**, 315–323 (1978).
9. Amtmann, E. & Sauer, G. *Nature* **296**, 675–677 (1982).
10. Kreig, P., Amtmann, E. & Sauer, G. *Analyt. Biochem.* **134**, 288–295 (1983).
11. Kreig, P. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 6446–6450 (1981).
12. Amtmann, E. & Sauer, G. *J. Virol.* **43**, 59–66 (1982).

© 1984 Nature Publishing Group

A protein from human placental nuclei binds preferentially to 5-methylcytosine-rich DNA

Lan-Hsiang Huang, Richard Wang, Miguel A. Gama-Sosa, Suresh Shenoy & Melanie Ehrlich

Department of Biochemistry, Tulane Medical School, New Orleans, Louisiana 70112, USA

The methylation of vertebrate DNA at the 5-position of ~3–10% of its cytosine residues occurs in a sequence-specific and tissue-specific manner[1–3] and has been implicated in the control of transcription[1,4–7]. How these differences are established and how they mediate the initiation or maintenance of transcription are unknown. DNA methylation might also have other roles, such as modulating DNA replication, transposition, DNA repair or chromosome configuration[1,8,9]. These other roles suggested for DNA methylation would be consistent with the finding that tissue-specific differences in methylation of certain gene regions[5], highly repeated satellite DNA sequences[8–10] and whole genomes[2,3] often do not correlate with transcriptional activity. For DNA methylation to modulate the expression, maintenance or duplication of chromosomes, there should be effector macromolecules, presumably proteins, which specifically recognize 5-methylcytosine (m^5C) residues in DNA. We describe here the first identification of a mammalian protein that binds preferentially to m^5C-rich DNA sequences.

This DNA-binding protein, which we refer to as methylated DNA-binding protein (MDBP), was isolated from nuclei of human placenta. It was detected in fractionated nuclear extracts assayed for their relative binding to human DNA experimentally enriched in m^5C and to the analogous DNA of a naturally low m^5C content[3]. These ligands were prepared by extensive nick-translation[11] of placental DNA with a radioactive mixture of the four standard deoxynucleoside triphosphates to give m^5C-deficient (~0.5 mol% m^5C) DNA, or with m^5dCTP replacing dCTP to yield m^5C-rich (9 mol% m^5C) DNA. Binding of proteins to ^3H-DNA or ^{32}P-DNA was assessed by a nitrocellulose filter-binding assay[12] (Table 1).

A preparation of nuclei from full-term placenta was extracted as described in Fig. 1 legend. Subsequent chromatography on phosphocellulose and then hydroxyapatite columns yielded a peak of DNA-binding activity having a marked preference for m^5C-rich DNA (Fig. 1). The peak fractions were dialysed against 25 mM NaCl, 50 mM Tris HCl *p*H 7.6, 0.1 mM Na_2EDTA, 0.2 mM dithiothreitol, 10% glycerol and then chromatographed on a DEAE-cellulose column using a linear gradient of 0.025 M to 0.40 M NaCl in the same buffer. The MDBP activity eluted as a discrete peak from 0.05 to 0.08 M NaCl. The proteinase K-sensitive MDBP preparation from this peak was stored in 50% glycerol and 200 μg ml^{-1} of acid-treated and neutralized[13] bovine serum albumin at −20 °C. In three placental isolations, MDBP activity eluted in the same fraction of the hydroxyapatite column and bound ~2–3 times more m^5C-rich than m^5C-deficient nick-translated human DNA. Under the same assay conditions, this ratio for the MDBP fraction from the DEAE-cellulose column was 3.5–4.5.

The specificity of MDBP for m^5C-rich DNA was confirmed with several DNAs. The same preferential binding to methylated DNA was observed with nick-translated human DNA in double-label experiments using m^5C-rich ^3H-DNA and m^5C-deficient ^{32}P-DNA as when the radiolabels on the DNAs were reversed (Table 1). All of these DNAs had been nick-translated until 35–50% of their nucleotides were replaced and so contained regions methylated on only one strand (hemimethylated), on both strands or on neither strand. To analyse the binding of

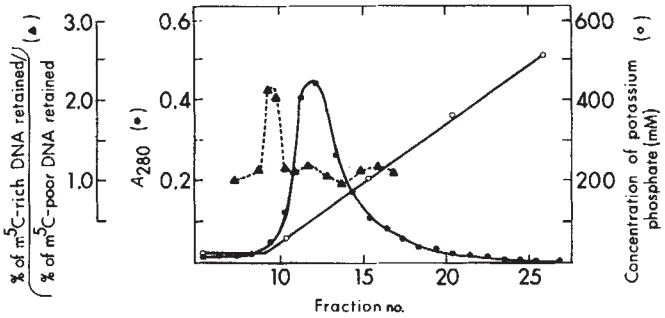


Fig. 1 Isolation of a human protein that binds preferentially to methylated DNA by hydroxyapatite chromatography. The crude nuclear fraction was prepared by homogenizing 550 g of placental trophoblast in 50 mM Tris-HCl *p*H 7.5, 1 mM $MgCl_2$, 0.25 M sucrose, 0.1 mM phenylmethylsulphonyl fluoride (PMSF), and 1 μg ml^{-1} of chymostatin, then centrifuging at 600 *g* for 10 min, and washing the pellet twice in 10 mM Tris-HCl *p*H 8.0, 0.3 mM $MgCl_2$, 0.1 mM PMSF, 1 μg ml^{-1} chymostatin. All steps in the purification were carried out at 0–5 °C. The pellet was extracted for 1 h in a final volume of 180 ml with 0.3 M NaCl, 10 mM Tris HCl *p*H 8.0, 1 mM Na_2 EDTA, 0.2 mM dithiothreitol, 10% glycerol, centrifuged at 30,000 *g* for 15 min, and the supernatant recentrifuged at 150,000 *g* for 3 h. This supernatant was diluted with 5 volumes of 25 mM potassium phosphate *p*H 7.4, 0.1 mM Na_2 EDTA, 0.2 mM dithiothreitol, 10% glycerol, and chromatographed on a phosphocellulose column (5×5 cm) with a linear gradient of 50 mM–1.5 M KCl in the above buffer (600 ml total volume). Slightly more binding to m^5C-rich DNA than to m^5C-deficient DNA was seen in the 0.12–0.22 M KCl fractions which were pooled and dialysed against the equilibration buffer (20 mM potassium phosphate *p*H 7.0, 0.1 mM Na_2EDTA, 0.2 mM dithiothreitol, 10% glycerol) for the subsequent hydroxyapatite column (2.5×4.5 cm). The sample was applied and eluted with a 200 ml gradient of this buffer and the same buffer containing 0.8 M potassium phosphate; 4.3-ml fractions were collected. Aliquots of 0.5–10 μl were incubated as described in Table 1 with a mixture of 10–20 ng of two nucleease S_1-treated, nick-translated human DNA preparations, one of which was enriched in m^5C (^{32}P-DNA), the other having a low level of m^5C (^3H-DNA). The relative amounts of ^{32}P- and ^3H-DNA retained after filtration through nitrocellulose filters was determined. The same single peak of MDBP activity was seen in analogous hydroxyapatite chromatograms after phosphocellulose chromatography of two other placental extracts. The MDBP activity used in the experiments of Tables 1 and 2 came from the DEAE-cellulose purification step (as described in the text) subsequent to hydroxyapatite chromatography. From the DEAE-cellulose column, 440 U of MDBP activity (1 U binds 10 ng of ^3H-labelled XP12 DNA in the presence of a 10-fold excess of unlabelled human DNA) and <200 μg of protein were obtained.

MDBP to completely hemimethylated DNA, replicative form (RF) bacteriophage M13 DNA (M13mp7-β-globin recombinant RFIV; refs 14–16) was prepared by extension of an oligonucleotide primer on a viral strand template in the presence of m^5dCTP. The ^3H-labelled hemimethylated M13 RF bound to MDBP approximately three times better than ^{32}P-labelled, nonmethylated M13 RF (Table 1). Even partially substituted, hemimethylated M13 RF which had only 9% of its cytosine residues methylated was bound better by MDBP than analogous nonmethylated DNA (Table 1).

The MDBP activity also showed a preference for a naturally m^5C-rich DNA, bacteriophage XP12 DNA (34 mol% m^5C, <1 mol% C; refs 17 and 18), compared with m^5C-deficient DNA. Bacteriophage XP12 DNA was used as a source of DNA with methylated C residues on both strands in competitive DNA binding assays in which unlabelled XP12 or human DNAs were competed against radiolabelled XP12 DNA or m^5C-enriched human DNA. The unlabelled XP12 DNA competed much

© 1984 Nature Publishing Group

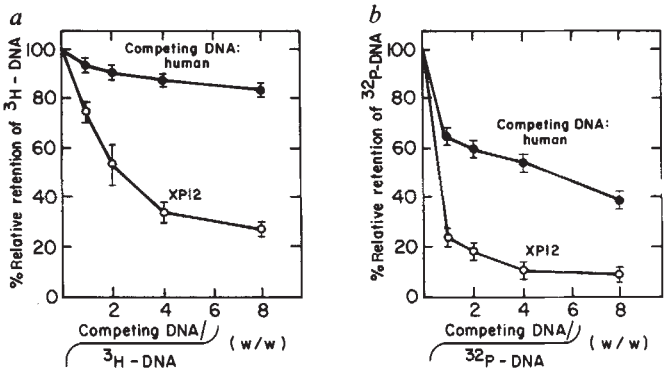


Fig. 2 Relative binding of the MDBP fraction to m^5C-rich XP12 DNA (*a*) and normal human DNA (*b*). MDBP from the DEAE-cellulose fraction was incubated with the indicated DNA samples. Protein binding to DNA was measured by quantitating the retention of DNA on nitrocellulose membranes under standard assay conditions in competitive DNA-binding assays with lightly sheared (10–20 kb), unlabelled human placental DNA or unlabelled XP12 DNA competing with 20 ng of ^3H-labelled, m^5C-rich XP12 DNA (*a*), or 10 ng of nick-translated, ^{32}P-labelled m^5C-enriched human placental DNA (*b*). The ability of unlabelled XP12 DNA to compete with labelled XP12 DNA was shown in other experiments to be unaffected by a decrease in length from 20 to 10 kb. The points shown are the average of duplicates with the ranges indicated. The percentages of cytosine residues methylated in these DNAs were as follows: XP12 DNA, >97%; human placental DNA, 3.6%; m^5C-enriched human placental DNA, 48%[3,17,18].

better than human DNA or phage N7 DNA (36 mol% C and <1 mol% m^5C) for binding to both radiolabelled m^5C-rich DNAs (Fig. 2 and Table 2). This specificity for m^5C-rich DNA was lost when the competing DNA was previously denatured or if the radiolabelled DNA in the assay was deficient in m^5C (Table 2). The latter result suggests that the MDBP preparation was contaminated with nonspecific DNA-binding proteins, and so we may be underestimating the specificity of MDBP for methylated DNA. The variety of m^5C-rich ligands that are bound preferentially by MDBP, and the absence of detectable Z-DNA structure[19] in naturally m^5C-rich XP12 DNA suggests that MDBP is not recognizing Z-DNA structures even though cytosine methylation predisposes runs of C-G dinucleotide sequences to formation of such structures[20,21].

It appears that MDBP is responding to methyl groups specifically at the 5-position of cytosine residues, since DNAs having higher contents of thymine, which contains the same 5-methyl group as in m^5C, were not bound preferentially by MDBP. These comparisons (data not shown) included bacteriophage λ DNA (25 mol% each T and C), bacteriophage N7 DNA (14 mol% T, 36 mol% C) and bacteriophage PBS2 DNA (36 mol% U, 14 mol% C). Nonglucosylased bacteriophage T4 DNA, which has all of its cytosine residues replaced by 5-hydroxymethylcytosine[22], competed no better against radiolabelled XP12 DNA than did human DNA for binding to MDBP (Table 2).

A role for the methylated DNA binding activity of MDBP *in vivo* is suggested by the specificity of MDBP for cytosine-methylated double-stranded DNA, by the fact that MDBP constitutes only a small proportion of the total DNA-binding protein and by its occurrence in the non-histone fraction of chromosomal proteins. Our final MDBP fraction had no DNA methyltransferase activity and had been well separated by hydroxyapatite chromatography from the main peak of DNA methyltransferase (data not shown). It also contained no detectable endonuclease or exonuclease activity. We suggest that MDBP binds to methylated DNA sequences *in vivo* as a regulatory protein that controls the template activity of DNA or as a structural protein that alters the local conformation of chromosomes. Studies of the nature and distribution of this protein and its use as a tool for detecting or obtaining methylated DNA sequences should further our understanding of the importance of DNA methylation in vertebrate cells.

We thank Dr Timothy Ley for the M13mp7-β-globin recombinant and Dr Stanley Hattman for the nonglucosylated T4 DNA. This research was supported in part by USPHS grant CA-19942 and a grant from the Louisiana Board of Regents.

Table 1 Binding of MDBP to methylated and unmethylated DNA

^3H-DNA	$m^5C/[C+m^5C]$	^{32}P-DNA	$m^5C/[C+m^5C]$	% Of ^3H-DNA retained/% of ^{32}P-DNA retained
Human	0.48	Human	0.02	3.6
Human	0.02	Human	0.35	0.35
M13 RF	0	M13 RF	0	1.0
M13 RF	0.09	M13 RF	0	1.3
M13 RF	0.25	M13 RF	0	2.0
M13 RF	0.50	M13 RF	0	3.3

The indicated combinations of ^3H-DNA plus ^{32}P-DNA were incubated in 0.1 ml of 10 mM sodium phosphate *p*H 6.8, 3 mM $MgCl_2$, 0.25 mM Na_2EDTA, 25 mM NaCl, and 20 μg ml^{-1} of acid-treated and neutralized[13] bovine serum albumin (BSA) at 25 °C for 15 min with 0.1 U (for M13 RF) or 0.5 U (for the human DNA) of MDBP from the DEAE-cellulose fraction. The mixtures were then diluted with 1 ml of the above buffer without BSA, and immediately filtered through nitrocellulose membranes[12]. The membrane filters were then washed three times with the dilution buffer. The human samples (20 ng for each assay) were placental DNA radiolabelled by nick translation with m^5dCTP (to give m^5C-rich DNA) or dCTP (for production of an m^5C-poor control) in the triphosphate mixture. To cleave at nicks and remove any single-stranded regions, these DNAs were treated with 80 U of S_1 nuclease per μg of DNA for 30 min at 37 °C in standard conditions[23] and then purified by phenol and chloroform extractions and ethanol precipitation to yield DNA fragments of ~1.5–2.5 kilobase pairs (kb). The M13 samples (2 ng for each assay) were relaxed. covalently closed circular DNA (RFIV; ref. 16) from M13mp7-β-globin (Probe F; ref. 15) radiolabelled by extension of a pentadecanucleotide primer with m^5dCTP or dCTP in various relative amounts in the triphosphate mixture, in a reaction catalysed by *Escherichia coli* DNA polymerase I and T4 DNA ligase. The resulting hemimethylated or nonmethylated RF was subsequently purified by denaturation and nitrocellulose filtration (S.S. and M.E., in preparation). All DNA samples were tested in duplicate and the average relative binding is given.

Table 2 Competition of various DNAs for binding to MDBP

Unlabelled competing DNA*		Radiolabelled DNA	
Type	$m^5C/[C+m^5C]$	% Competition with ^3H-XP12 DNA†	% Competition with m^5C-poor, ^3H-human DNA
XP12	>0.97	77	30
Human	0.03	25	29
N7	<0.01	14	12
T4	<0.01	25	39
Single-stranded XP12	>0.97	19	—
Single-stranded human	0.03	19	—
Single-stranded N7	<0.01	18	—

* In this experiment, 20 ng of radiolabelled DNA and 160 ng of unlabelled DNA which had been sheared to a molecular length of ~10–20 kb, were assayed in duplicate for binding to 0.6 U of MDBP in standard conditions. Phage XP12 DNA, which is naturally rich in m^5C in both strands, was labelled *in vivo* with [6-^3H]uracil[24] and DNA from diploid human fibroblasts was labelled in cell culture with [^3H-methyl]thymidine. Phage N7 DNA has a low A+T content like XP12 DNA but N7 DNA contains <1 mol% m^5C. Single-stranded DNA was obtained by heat-denaturation and quick cooling of native DNA. Phage T4 DNA naturally contains glucosylated 5-hydroxymethylcytosine residues replacing all of the cytosine residues. The T4 DNA used in these assays was from a mutant which contains nonglucosylated 5-hydroxymethylcytosine residues instead of glucosylated residues. The unlabelled human DNA was unsubstituted placental DNA, which has 0.76 mol% m^5C[3].

† The per cent competition is defined as the per cent retention of the radiolabelled DNA in the absence of competitor (R_o) minus the per cent retention in the presence of competitor (R_c) times 100 and divided by R_o: $100 \times (R_o - R_c)/R_o$. The averages from duplicate samples in the standard nitrocellulose filter-binding assay were used to determine these values.

© 1984 Nature Publishing Group

Received 15 September 1983; accepted 9 January 1984.

1. Ehrlich, M. & Wang, R. Y.-H *Science* **212**, 1350–1357 (1981).
2. Gama-Sosa, M. A. *et al. Biochim. biophys. Acta* **740**, 212–219 (1983).
3. Ehrlich, M. *et al. Nucleic Acids Res.* **10**, 2709–2721 (1982).
4. Felsenfeld, G. & McGhee, J. *Nature* **296**, 602–603 (1983).
5. Cooper, D. N. *Human Genet.* **64**, 315–333 (1983).
6. Doerfler, W. *A. Rev. Biochem.* **52**, 93–124 (1983).
7. Busslinger, M., Hurst, J. & Flavell, R. A. *Cell* **34**, 197–206 (1983).
8. Sano, H. & Sager, R. *Proc. natn. Acad. Sci. U.S.A.* **79**, 3584–3588 (1982).
9. Gama-Sosa, M. A., Wang, R. Y.-H., Kuo, K. C., Gehrke, C. W. & Ehrlich, M. *Nucleic Acids Res.* **11**, 3087–3095 (1983).
10. Sturm, K. S. & Taylor, J. H. *Nucleic Acids Res.* **9**, 4536–4546 (1981).
11. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. *J. molec. Biol.* **113**, 237–251 (1977).
12. Riggs, A. O., Suzuki, H. & Bourgeois, S. *J. molec. Biol.* **48**, 67–83 (1970).
13. Lacks, S. & Greenberg, B. *J. biol. Chem.* **242**, 3108–3120 (1962).
14. Messing, J. & Viera, J. *Gene* **19**, 269–276 (1982).
15. Ley, T. J., Anagnou, N. P., Pepe, G. & Nienhuis, A. W. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4775–4779 (1982).
16. Baas, P. D. *et al. J. molec. Biol.* **152**, 615–639 (1981).
17. Kuo, T.-T., Huang, T.-C. & Teng, M.-H *J. molec. Biol.* **34**, 373–375 (1968).
18. Kuo, K. C., McCune, R. A., Gehrke, C. W., Midgett, R. & Ehrlich, M. *Nucleic Acids Res.* **8**, 4763–4776 (1980).
19. Huang, L.-H., Farnet, C. M., Ehrlich, K. C. & Ehrlich, M. *Nucleic Acids Res.* **10**, 1579–1591 (1982).
20. Klysik, J., Stirdivant, S. M., Singleton, C. K., Zacharias, W. & Wells, R. D. *J. molec. Biol.* **168**, 51–71 (1983).
21. Behe, M. & Felsenfeld, G. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4597–4601 (1981).
22. Warren, R. A. J. *A. Rev. Microbiol.* **34**, 137–158 (1980).
23. Vogt, V. M. *Eur. J. Biochem.* **33**, 192–200 (1973).
24. Ehrlich, M., Ehrlich, K. & Mayo, J. A. *Biochem. biophys. Acta* **395**, 109–119 (1975).

Fish antifreeze protein and the freezing and recrystallization of ice

Charles A. Knight*, Arthur L. DeVries† & Larry D. Oolman*

* National Center for Atmospheric Research, Boulder, Colorado 80307, USA
† Department of Physiology and Biophysics, University of Illinois, Urbana, Illinois 61801, USA

Antifreeze glycopeptides and peptides from the blood of polar fishes prevent the growth of ice crystals in water at temperatures down to ~1 °C below freezing point, but do not appreciably influence the equilibrium freezing point[1–6]. This freezing point hysteresis must be a disequilibrium effect, or it would violate Gibbs' phase rule, but the separate freezing and melting points are experimentally very definite: ice neither melts nor freezes preceptibly within the 'hysteresis gap', for periods of hours or days[4,7]. We report here unusual crystal faces on ice crystals grown from solutions of very low concentrations of the antifreeze glycopeptides and peptides. This is a clue to the mechanism of freezing inhibition, and it may be the basis of a simple, very sensitive test for antifreeze material. Very low concentrations of the antifreeze protein are also remarkably effective in preventing the recrystallization of ice.

Estimates of the equilibrium freezing points of solutions of antifreeze proteins were obtained by measuring the vapour pressure of solutions with a Wescor vapour pressure osmometer, which uses a small disk of filter paper soaked with a drop of solution. This showed that the equilibrium freezing points of the solutions are virtually equal to the measured melting points. We duplicated this fundamental result directly, by using a known amount of glycopeptide dispersed within a thin glass frit, then weighing the frit after it reached equilibrium with ice at a constant temperature in an atmosphere of pure water vapour. Thus, the freezing hysteresis represents the inhibition of freezing with little or no effect on melting, and the mechanism of this inhibition is the problem in which we are interested.

Eight glycopeptides having this special antifreeze property (antifreeze glycoproteins, AFGPs) have been isolated from Antarctic fishes[8]. They range in molecular weight (MW) from 2,600 to 33,700, and are polymers of a repeating sequence of Ala-Ala-Thr with the disaccharide galactose-*N*-acetylgalactosamine linked to each threonine[9]. Freezing point depressing activity depends on chain length, with the high molecular weight glycopeptides more effective in depressing the freezing point than those of low molecular weight; see Fig. 1. In our experiments we used a mixture of the five heaviest (1–5) glycopeptides because they are easiest to separate and contribute the most to the hysteresis, and pure 8, the lightest. Some temperate zone fish also produce antifreeze peptides in the winter[8,10–12], and one of these was also tested.

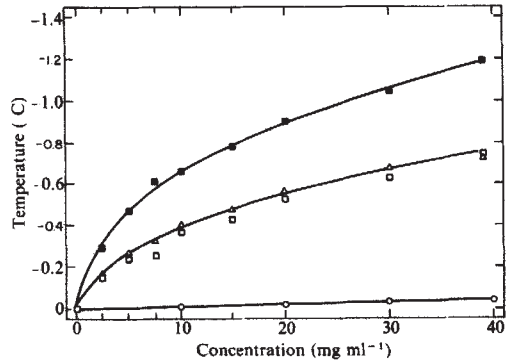


Fig. 1 Freezing points of aqueous solution of high molecular weight glycopeptides 1–5 (■), and of low molecular weight glycopeptides 7 (△) and 8 (□) as a function of concentration. The melting temperatures (○) are the same for all three solutions.

During freezing, AFGPs are incorporated into growing ice and not ploughed ahead of the advancing interface like normal solutes[7]. Ice growth in concentrated solutions of antifreeze is in the form of spicules, or fibres, parallel to the *c*-axis, the hexagonal symmetry axis, of the ice[7]. X-ray diffraction studies of ice grown from such solutions show only normal ice, thus there is no new phase involved[7].

A simple, temperature-gradient microscope stage was constructed[13] and the freezing of 300-μm layers of AFGP solutions between glass microscope slides was studied. The apparatus allows detailed observation of the freezing interface either from slightly supercooled solutions (one supercools the liquid and then nucleates ice at one end) or by establishing a stationary ice–solution interface using the temperature gradient, and then moving the slide so that the interface melts back or moves forward.

For the mixture of AFGPs 1–5, the hysteresis effect was confirmed with the new apparatus, and at concentrations of 1% (by weight) or more, where the hysteresis is large, the fibrous structure was evident (Fig. 2*a*). As ice growing from pure water at low velocities is not fibrous and grows slowest parallel to the *c*-axis (Fig. 2*b*) it was of interest to observe the growth habit transition as a function of the concentration of AFGP. Small concentrations, from $\sim 10^{-4}\,g\,g^{-1}$ down to $\sim 10^{-7}\,g\,g^{-1}$, or $\sim 10^{-7}$ to 10^{-10} M, bring about the expression of other faces (Fig. 2*c*) which we call $\{10\bar{1}x\}$. Accurate angle measurements have not been possible, but the faces are sometimes curved in such a way that the last index is continuously variable. With increasing AFGP concentration, the new faces cease to be evident: the fastest growth direction becomes parallel to the *c*-axis, the growth units become smaller and smaller, and eventually, at 0.5–1% (by weight), the growth becomes fibrous.

This is the only report of crystal faces other than {0001} on ice growing from liquid water or solutions, except for one occurrence in a water-filled crevasse within a temperate glacier, where organic impurities were a possible cause[14]. The profound influence of AFGP on growth habit of ice at very low AFGP concentrations confirms the activity of the AFGP molecules at the ice–water interface and their interference with the crystal growth mechanism, and suggests that they do act by inhibiting growth perpendicular to the *c*-axis. Several organic compounds of comparable or higher molecular weight that do not produce

© 1984 Nature Publishing Group

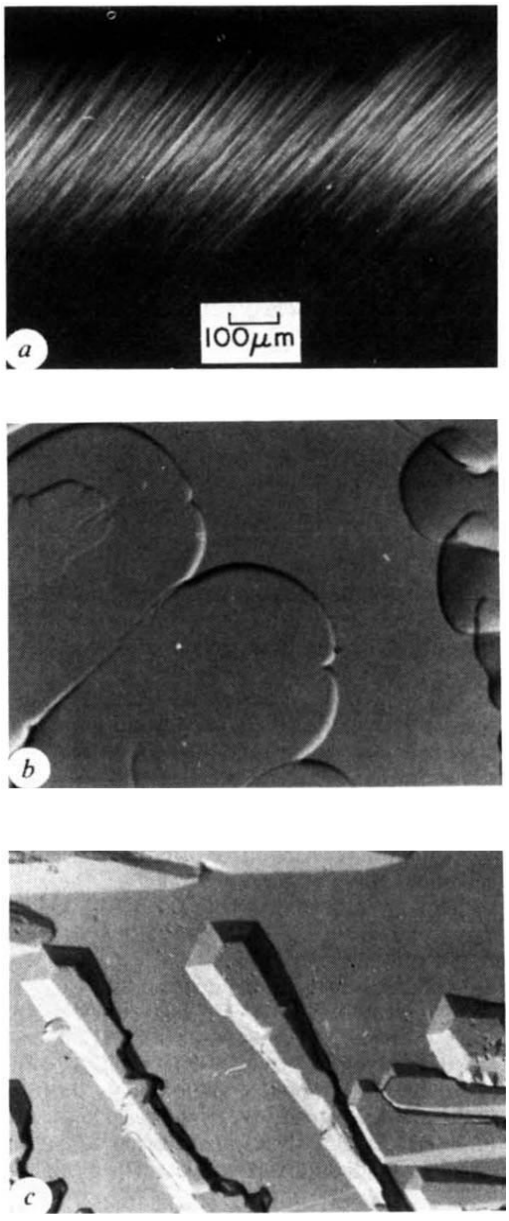


Fig. 2 Effect of the glycopeptides on crystal growth habit. *a*, The freezing interface of a 1% (by weight) solution of glycopeptides 1–5; water above, ice below. The fibres are parallel to the *c*-axis of the ice. *b*, Freezing of pure water at low supercooling. The *c*-axis is perpendicular to the plane of the photograph, and all interfaces are rounded except the basal face. Growth is slowest parallel to the *c*-axis. *c*, On freezing of a 2.5×10^{-5} g g^{-1} solution of glycopeptides 1–5 ($\sim2\times10^{-8}$ M) at low supercooling, $\{10\bar{1}x\}$ faces develop. The *c*-axis, again, is perpendicular to the plane of the photograph. The scale in *a* applies to all photographs.

freezing hysteresis [polyethylene glycol (MW 6,000), dextran (MW 9,400 and 20,000) and polyvinyl pyrrolidone (MW 40,000)] were tested and found not to produce growth habit modification. A peptide antifreeze isolated from winter flounder blood[8] that also produces freezing hysteresis, has a similar effect on ice growth habit at low concentrations. This habit modification effect might be the basis of a very simple and sensitive test for activity of a potential antifreeze material, as the amount needed to produce and detect the effect is very small.

We have also found that concentrations of AFGP as low as 10^{-9} g g^{-1} ($\sim10^{-12}$ M) are extremely effective in preventing recrystallization of ice, whereas the same organic compounds as noted above were found to have little effect on recrystallization. This strong effect of very small amounts of AFGPs on recrystallization might have application in the preservation of frozen foods or medical materials[15]. Although the importance of the recrystallization process *per se* has not been firmly established, it has been suggested as a possible role of the AFGPs in the survival of some freeze-tolerant insects (J. Duman, private communication).

The mechanism by which the AFGP inhibits freezing is not understood. Certainly it bonds at the ice–water interface, and it has been suggested that it might inhibit the lateral growth of steps[4,6,8]. If this were the case, it is not clear why low concentrations stimulate the development of the crystal faces that are observed. Jackson[16] calculated that the equilibrium structures of all surfaces of ice should be molecularly rough except the basal face, and this fits the observation that only basal faces appear in ice growth from pure water or, in fact, from any water solutions of which we are aware except those of the antifreeze proteins. An explanation of why the low amounts of AFGP produce the new crystal faces may be a key to understanding the freezing hysteresis itself.

Strong effects of small amounts of impurities on the growth habit of crystals grown from solution or from vapour are common. However, this is the only example of which we are aware, of such a striking impurity effect in crystal growth from the melt. It is also, as far as we know, the only example of freezing point hysteresis induced by a molecular solute. Some polymeric solutes that form gels are known to induce a similar freezing hysteresis[17].

This work was supported in part by NSF grant DPP 81-16917 (to A.L.D.). The National Center for Atmospheric Research is sponsored by the NSF.

Received 15 August; accepted 22 December 1983.

1. DeVries, A. L. *Science* **172,** 1152–1155 (1971).
2. DeVries, A. L. in *Biochemical and Biophysical Perspectives in Marine Biology* Vol. 1 (eds Sargent, J. S. & Mallins, D. W.) 289–330 (Academic, London, 1974).
3. Feeney, R. E. *Am. Scient.* **62,** 712–719 (1974).
4. DeVries, A. L. & Lin, Y. in *Adaptation Within Antarctic Ecosystems*, 439–458 (Gulf Publ. Co., Houston, 1977).
5. Feeney, R. E. & Yeh, Y. *Adv. Protein Chem.* **32,** 191–282 (1978).
6. DeVries, A. L. in *Animals and Environmental Fitness* (ed. Gilles, R.) 583–607 (Pergamon, Oxford, 1980).
7. Raymond, J. A. & DeVries, A. L. *Proc. natn. Acad. Sci. U.S.A.* **74,** 2589–2593 (1977).
8. DeVries, A. L. *Comp. Biochem. Physiol.* **73A,** 627–640 (1982).
9. Schrag, J. D. & DeVries, A. L. *Comp. Biochem. Physiol.* **74A,** 381–385 (1983).
10. Hew, C. L., Fletcher, G. L. & Ananthanarayanan, V. S. *Can. J. Biochem.* **58,** 377–383 (1980).
11. Ananthanarayanan, V. S. & Hew, C. L. in *Biomolecular Structure, Conformation, Function and Evolution* Vol. 2 (ed. Srinivasan, R.) 191–198 (Pergamon, Oxford, 1980).
12. Slaughter, D., Fletcher, G. L., Ananthanarayanan, V. S. & Hew, C. L. *J. biol. Chem.* **256,** 2022–2026 (1981).
13. Jones, D. R. H. *Metals Mater.* **6,** 312–313 (1972).
14. Knight, C. A. & LaChappele, E. R. *J. Glaciol.* **9,** 87–101 (1970).
15. Mazur, P. *Science* **168,** 939 949 (1970).
16. Jackson, K. A. in *Liquid Metals and Solidification*, 174–186 (Am. Soc. for Metals, Cleveland, 1958).
17. Kuhn, W. *Helv. chim. Acta* **39,** 1071–1086 (1956).

Nature in Tokyo

***Nature* has opened an office in Tokyo. Dr Alun Anderson, for the past few years in charge of the News and Views section of *Nature*, will be in charge of the office for an initial period of one year. By this means, *Nature* hopes more efficiently to cover news of developments in Japan and also to provide tangible help to Japanese scientists wishing to submit research papers or other communications to *Nature*.**

Dr Anderson's address is *Nature*, Macmillan Shuppan KK, Eikow Building, 10-9, Hongo 1-chome, Bunkyo-ku, Tokyo 113 (Telephone (03) 816 3756/7; Telex JATOKJ32121 MSKK). □

© 1984 Nature Publishing Group

Telling stories

Bryan Silcock

Presenting Science to the Public. By Barbara Gastel.
ISI Press, Philadelphia: 1983. Pp.146. Hbk $17.95 (North America), $20.95 (elsewhere); pbk $11.95 (North America), $14.95 (elsewhere).

JUST over 30 years ago *Nature* published a report of what is now recognized as one of the key discoveries of modern biology. Like most scientific papers it was written in a kind of code. Who outside the small group then interested in DNA could have recognized the significance of Watson and Crick's structure from their letter alone? The only hint that it was anything out of the ordinary was the throwaway line "It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material", and this was intended not to inform but to establish a claim.

Yet the structure of DNA was clearly of immense general interest. It answered a fundamental question about life that occurs to anyone with a little intellectual curiosity, and, as we now know, it had far-reaching implications for medicine and technology. In 1953 they could not be foreseen in detail, but genetic manipulation was obviously among them.

The formal methods of scientific communication are, in fact, of little use to people who want a broad view of what is going on in science: scientists interested in disciplines other than their own; politicians who vote funds for research; taxpayers who provide them. Science *does* need to be presented to the public. Barbara Gastel, who teaches science journalism at the Massachusetts Institute of Technology, takes most of this for granted. Her book is very much a practical manual, for the working scientist wishing to try his hand at a bit of popularization and for his colleague who willy-nilly becomes the object of media attention. Nearly half of it is concerned with presenting science through the mass media.

The advice offered is well informed, sensible and concise, though the occasional passage such as "the main item that succeeds interaction with a journalist is likely to be a story" does seem to have diffused into the main text from some list of things writers should not do. However, as a participant in many such interactions I can vouch for her understanding of the problems. Any scientist unfamiliar with the workings of the media will find the book useful.

It is said to be a good working rule that relations between journalists and politicians should be bad. I do not believe the same rule applies to journalists and scientists, but it must be admitted that relations are often strained. Why is there so much friction? Leaving aside the occasions on which reporters simply get things wrong, it does seem from where I stand, on the media side of the fence, that those on the other side must take some of the blame for a lot of unnecessary misunderstanding. Many scientists seem to regard a press report on their work as a kind of debased research paper. It is of course nothing of the kind. As the example of DNA shows, the scientific paper is a very formalized mode of communication; it makes no attempt to put the work described in any kind of broad perspective, and deliberately eschews extrapolation and speculation. Yet for the general public the first is essential, the second desirable and the third often useful. Governments fund research and taxpayers pay for it because they believe it will bring benefits. They are entitled to know both what is just around the corner and what is looming over the horizon.

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

A SUPERJAB that will cure certain types of cancer will be available in Britain within a year.

The amazing breakthrough was revealed yesterday by pioneering British scientist Dr Mike Waterfield

Journalism at its worst — but do scientists appreciate the demands of the media?

The simplification that accompanies popularization is another common source of friction. Although nearly all scientists concede that complex ideas need to be simplified for a general audience, they are often unhappy with the results. Many of the things they object to are the unavoidable consequences of the simplification and compression essential in the popularization of science.

As Gastel says, a good working rule is never to underestimate people's intelligence and never to overestimate their knowledge. But consider the application of this rule to, say, a popular account of the recent discovery of the W and Z particles at CERN. The line such an account should take is clear enough: the particles were predicted by a theory that unifies two of the four forces in nature; the discovery of the W and Z therefore shows that the theory is on the right lines. But most people do not know there are four natural forces, let alone what they are. They are probably pretty hazy about fundamental particles, atoms and nuclei. A newspaper or TV report cannot afford to have a didactic air if it is to command any attention, and in any case there is not the space for long explanations. So reporters are forced to resort to fudges, short cuts and minor inaccuracies; calling particle accelerators atom smashers, for example, because they think their readers probably have not heard of particle accelerators but that they might have heard of atom smashers.

Gastel's description of these and other problems should help to provide the understanding that leads to tolerance. It is unfortunate though that she has virtually nothing to say about the most important aspect of the presentation of science to the public today: the way the media handle issues such as nuclear energy and genetic engineering that are at once technically complex and politically or ethically sensitive. In the long run an inaccurate newspaper report about fundamental particles does not matter. What newspapers and television have to say about controversial issues does. For better or worse the media have played a major part in promoting the anti-nuclear movement, and a few years ago an alarmist British television programme about criteria for brain death led to a sharp fall in the supply of kidneys for transplant operations.

Media coverage of such issues is often appallingly bad. Perhaps this is partly because journalists like above all to see themselves as exposers of wrongs, and as the champions of lonely crusaders against the system. Such attitudes help to preserve democratic freedoms, but they are not conducive to the sober analysis of complex technical issues and statistical arguments. The facts can begin to seem part of the system. In defence, I can only point out that the balls so enthusiastically kept in play are nearly always first thrown onto the field by a member of the scientific community. This is true of even the daftest ideas. The warning that evil governments might perpetuate themselves by cloning millions of compliant citizens was, if I remember rightly, first issued by a very eminent biologist.

So what should a scientist do when he is dragged into a media debate over whose form and content he has no control? Gastel points out in another context the importance of being patient and of not being patronizing — she might also have mentioned that there is nothing that so arouses the suspicions of a journalist as the feeling that information is being withheld. □

Bryan Silcock is science correspondent of the Sunday Times.

© 1984 Nature Publishing Group

Putting it all together again

R.H. Pritchard

Growth of the Bacterial Cell.
By J.L. Ingraham, O. Maaloe and F.C. Neidhardt.
Sinauer/Blackwell Scientific: 1983. Pp.435. $25, £19.75.

THE end-point of cell biology is a description of living cells as a unity. But the use of mutants to identify and analyse the bits from which cells are constructed has been so rewarding that efforts to understand how they work in concert have taken second place. This book thus joins a select company of texts on bacterial physiology; it attempts to describe the structure and function of bacterial cells in the context of their growth and division.

The approach chosen is first to examine the structure and composition of a typical bacterial cell using *E. coli* as a model. In fact, the book could just as well have been given the title "Growth of *E.coli*". Later sections look at the assembly of major cell components (such as the envelope and ribosomes); the polymerization of key macromolecules (DNA, RNA, protein, peptidoglycan, phospholipid, polysaccharides); fuelling reactions; and transport. There is a long section devoted to growth of individual cells and of cultures, dealing with concepts such as balanced growth, with the theory of continuous culture, and with cell composition and growth yield under different cultural conditions. Control of gene activity is also covered in detail.

The book ends with an attempt to provide an analysis of the variation of cell composition with growth rate, couched in terms of models which assume that composition is optimized to achieve maximum rate and efficiency.

Within the boundaries set for themselves, the authors' style is refreshingly clear and their treatment thorough. Even well-worn topics such as the operon model of control are made interesting. Speculation about the adaptive significance of the phenomena described provokes the reader's attention and only occasionally gets out of hand ("It is the inclusion of the entire genome in a single molecule that makes the prokaryotic mode of DNA replication and cell division workable").

Nevertheless, the coverage of some topics is disappointingly slight or even absent. One example is the treatment of the interrelated subjects of growth of the envelope, cell shape and septation. Another is initiation of DNA synthesis and its control. Recent discoveries about the control of plasmid replication surely deserved a mention; while in view of its key role in cell growth, the control of initiation of chromosome replication also merited more serious treatment. The cursory coverage of this topic perpetuates such misconceptions as the belief that there must be proteins involved in initiation that are consumed in the process.

Are not osmoregulation and control of *p*H also central topics for a book on growth physiology? They are not mentioned. On the other hand, the extensive treatment of bacterial genetics seems out of place, as do a number of the appendices ("Genetic Mapping" and "A Genetic Approach to Characterising Complex Promotors", for example).

Readers familiar with the excellent book *Control of Macromolecular Synthesis* by O. Maaloe and N.O. Kjelgaard (W.A. Benjamin, 1966), may be disappointed that this new one does not take the analysis of cell growth much further forward. It is essentially a restatement of the old ideas. Maaloe's hypothesis of passive regulation — suggested as the mechanism which optimizes the proportion of protein synthesizing capacity which is devoted to the synthesis of the protein synthesizing system itself — is discussed at length although it remains largely speculative and does not seem easy to reconcile with the observation that the rate of protein synthesis is not affected by a reduction in DNA concentration.

Despite these shortcomings, the book does contain much good quantitative information about *E. coli* that is not readily found elsewhere. There are also helpful quantitative problems at the end of each chapter. The book assumes a knowledge of general biochemistry and metabolic pathways, and this allows the authors to offer a sophisticated text that biochemistry students can use to take their understanding of bacterial physiology forward. Those who think of themselves as geneticists or as molecular biologists will also find in it a new and useful perspective on cell growth. □

R.H. Pritchard is Professor of Genetics at the University of Leicester.

Grains of evidence

Peter D. Moore

An Atlas of Past and Present Pollen Maps of Europe 0-13000 Years Ago.
By B. Huntley and H.J.B. Birks.
Cambridge University Press: 1983. Pp.667. £85, $174.

QUATERNARY palynologists are, by nature, patient people; a thousand years in their sight is but a few centimetres of sediment. But the publication of the Huntley and Birks pollen maps has, nevertheless, been awaited with eager, sometimes almost frenzied expectation. The reason for this lies in the fact that these maps represent the digestion and collation of a vast bank of data abstracted from published information, usually pollen diagrams, from a total of 843 geographical locations scattered over Europe. All European palaeopalynologists and many from other parts of the world must have spent a large portion of their working lives poring over a mere fraction of these European diagrams and I am confident that no one has previously drawn information systematically from them all. The maps, therefore, offer us speedy and comfortable access to an abundance of data — hence the anticipation.

The main aim of the *Atlas* is to provide a graphic representation of changing vegetation on a continental scale over the past 13,000 years, that is from the latter part of the last glaciation. In the main, this is achieved by plotting pollen contours for each of the major pollen taxa at time intervals varying from 500 to 2,500 years, depending upon rates of change. For trees and shrubs the values plotted are expressed as a percentage of trees plus shrubs, and for herbs as a percentage of total land pollen (plus the taxon under consideration if it was omitted from the original pollen sum). Pollen taxonomy is on the whole conservative, being limited by the degree of discrimination used by the many primary authors. Thus Ericaceae is not subdivided, though *Quercus* is separated into deciduous and evergreen types. A separate map is plotted for each taxon at each sampling interval and is preceded by a discussion of the taxonomy and pollen dispersal characteristics of that taxon, together with a description and interpretation of the main changes observed.

The selection of sites for data abstraction is clearly an important aspect of the research underlying this book. Mainly those which are well dated using radiocarbon techniques are used. This was obviously necessary, though it has resulted in the exclusion of some significant data, such as the contribution of *Phillyrea* to the pollen rain of southern Iberia. Ideally, only influx data would have been used but in practice this is impossible because of the lack of sites from which such information is available. The use of percentage data undoubtedly results in some confusing pictures, however; a tree species which is invading an open landscape, for example, would be better represented in proportional terms than one invading a region already bearing forest, though invasion rate and population density may be the same. Sites with old sediments (>10,000 years) are generally less frequent than those with recent material, so the resolution of the contour maps is less fine for these. Overlay sheets showing the precise location of the sites abstracted for each time sample are vital for the indication of this resolution capacity and are supplied, together with overlays of physical features and national boundaries.

© 1984 Nature Publishing Group

A series of present-day pollen maps is also provided, which is of use as a datum from which the past can be viewed. Also of interest is the principal components analysis and mapping. Here the total data set at each sampling interval is ordinated and the sites mapped according to their scores on each axis. Distinct patterns emerge and can be interpreted on the basis of the loadings of the taxa on each component. In this way the distribution of major vegetation formations at each sample time can be discerned and their changing patterns observed.

The outcome of these analyses and compilations is an excellently presented set of maps which amply rewards careful study. In them one can trace the glacial refugia of our now familiar tree species, and the routes by which they re-invaded central and northern Europe, previously laid bare by glacial and periglacial climates. One can observe the complex vegetation patterns of modern times emerging as a consequence of the interaction between deteriorating climate and the progressive management of the landscape by human beings.

The beech (*Fagus*), for example, is represented by just a few scattered pollen grains in the Balkans 10,000 years ago. By 6,500 it had managed to reach northern Italy and was squeezing around the east of the Alps into Austria. Then came a surge westwards into southern France, northward through Poland to the Baltic and finally the spread to the north-west and its entry into Britain probably between 3,000 and 4,000 years ago. Archaeologists will no doubt be comparing such movements with cultural developments in Neolithic and Bronze Age Europe to see if its ultimate spread can be linked with primary forest disturbance and secondary invasion.

Pine reached Britain much earlier; certainly by 10,000 years ago. One does, however, gain a distinct impression from these maps that the Scottish population had a separate origin from the English one, the latter being clearly part of a northward movement in Europe, whereas the former seems to be linked with Scandinavia. This should add fuel to the debate, generated by recent cytological work, over the taxonomic affinities of the Caledonian pines.

Presented in simple, graphic form, there is such a wealth of information here that all palynologists will find new insight into their own particular field of interest. The expectancy of past months will now be replaced by an industrious hush in the palynological world as we each work out the implications of these maps for our own pet theories and projects. It is a splendid book, well worth waiting for, and the perfect present for the armchair biogeographer. □

Peter D. Moore is Senior Lecturer in the Department of Plant Sciences, King's College, University of London.

Explorations in neurobiology

Charles F. Stevens

Membrane Potential-Dependent Ion Channels in Cell Membrane: Phylogenetic and Developmental Approaches.
By S. Hagiwara.
Raven: 1983. Pp.115. $29.50.

THE world has explorers, those who discover new lands, and map makers, those who chart out what the explorers have found. Although Hagiwara — universally known as Hagi in the trade — has made his share of maps, he is one of neurobiology's foremost explorers, and with this book has given us a journal, written retrospectively, of his travels. He writes in the first person, informally and clearly, and not only describes what he has discovered, but also what he was searching for, why, and what he thought about the things he turned up.

Hagiwara has had an important role in the development of neurobiology and membrane biophysics in many ways, but two of his general influences require special note. First, he has been a leader in informing neurobiologists of the value of a comparative approach. Not so many years ago, a prejudice was abroad among brain scientists that studies of lower organisms are largely irrelevant for revealing the mysteries of nervous systems such as our own. A few invertebrates — the squid, for example — were awarded honorary mammalhood, but mostly invertebrate and non-mammalian vertebrates were tacitly considered unworthy. Hagi, together with Kuffler and others, showed us that the systems that best reveal the properties of the brain can be invertebrate neurones or certain cells (such as eggs) that are not even neuronal.

To understand Hagiwara's second main area of influence, one must appreciate the situation in neurobiology about 1960. Intracellular recording was just starting to be widely used for a variety of different systems, and voltage clamp studies were rare. In those days all neuronal membranes were simple things, equipped with only the Hodgkin–Huxley sodium and potassium channels. We now know, of course, that the brain's electrical activity depends upon a great number of different channel types. This insight depended in large part on results of Hagiwara's studies described in his new book.

The book itself is not long, just over 100 pages which are divided into five main chapters, with a brief introduction and conclusion. These chapters focus on properties of voltage gated sodium, calcium and potassium channels, and the descriptions appropriately centre on Hagiwara's own work — most of the figures are, for example, from his papers — although other contributions are not neglected. The writing is straightforward and easy to understand, but this is not a book for the casual reader who wants to learn something about the importance of channels to our understanding of brain function. Those who study channel function, however, will want to read this illuminating, personal account of Hagiwara's contributions of the past quarter century. □

Charles F. Stevens is Chairman of the Section of Molecular Neurobiology at Yale University School of Medicine.

Chromosome cookery

Michael Ashburner

Working with Animal Chromosomes.
By Herbert Macgregor and Jennifer Varley.
Wiley: 1983. Pp.250. £17.50, $33.

RECIPE books have a long and distinguished history. Mrs Beeton's *Cookery and Household Management* has set many a married couple on the path to conjugal bliss, a path made smoother by good cooking. Macgregor and Varley now offer, to a rather narrower audience than Mrs B, *Working with Animal Chromosomes.* Their intention is "to describe in detail ways of working with chromosomes that are most useful for tackling current problems relating to the organization, function and behaviour of the genome in eukaryotic cells". Nearly half of the book is devoted to giant chromosomes — the polytene chromosomes of flies (but not, alas, of protozoa) and lampbrush chromosomes, especially those of the amphibian oocytes. This emphasis reflects the authors' own research interests and, of course, their aims.

The analogy between cooking and cytological techniques is a good one. Both demand a degree of flair and extensive practice. Both can be learnt from books, but should be learnt from a master. For both, some appreciation of the underlying chemical and physical principles is a help, but many of the best cooks, and cytologists, use intuition and "feel", rather than science, in their art. At the very best any recipe book can only inform the reader that some techniques are possible and that close attention to the recipe will produce a result that is broadly successful. To do more will require practice.

Macgregor and Varley have quite deliberately not attempted to be comprehensive nor have they included any serious consideration of the theoretical background to the methods they describe. I would support the selectivity of their coverage; once a basic technique has been learnt it is usually quite easy to pick up a

© 1984Nature Publishing Group

more specialized method based upon it. However, the price for selectivity is over-generalization, the risk of misleading the reader. By and large the authors have avoided this problem.

I am less happy about the lack of theory, particularly in the final chapters on *in situ* hybridization, autoradiography and the measurement of chromosomal DNA. I accept that the authors have provided extensive references to more technical papers. But one aim for a book of this type is that it should be reasonably self-contained. I would also like to have seen a fuller discussion of the pitfalls in the interpretation of data. To give but one example, a method is given for the *in situ* hybridization of DNA probes to chromosomal RNA. This technique is valuable, especially for transcripts from non-repetitive sequences (not considered by Macgregor and Varley) but must be well controlled against the artefact of DNA-DNA hybridization. The methods for controlling this technique are, unhappily, not included.

There will, however, be few cytologists who cannot read this book for profit. Macgregor and Varley have written a useful introduction to modern methods for working with animal chromosomes, and have done so in a most readable way. □

Michael Ashburner is Reader in Developmental Genetics at the University of Cambridge.

Language of physics

Nevill Mott

Reflexions d'un Physicien.
By Anatole Abragam.
Hermann, 293 rue Lecourbe, 75015 Paris: 1983. Pp.152. FF 58.

ANATOLE Abragam is one of the most distinguished of French theoretical physicists; he is now a professor at the Collège de France, and was this year elected as a foreign member of the Royal Society. His speciality has been research on nuclear magnetism. The atomic nucleus carries a small magnetic moment, which can be detected by resonance methods, and nuclear magnetic resonance (NMR) is now a useful tool in many branches of science.

This book contains lectures, articles and letters from throughout Abragam's career. A third of it is taken up with two lectures on nuclear magnetism, the first of which, given in 1960 at the Collège de France, tells amusingly how research on NMR started independently in Stanford and in Harvard, with different concepts and techniques. For years afterwards one could tell from which side of the continent a researcher came "plus surement que par son accent". The second continues the story from 1960 to the present day, giving a very readable account of the present position of his subject. "Magnus" he calls it, rejecting "Magnuc", a current abbreviation, because of its ugliness.

Abragam is sensitive to the way things are written. He claims that George Orwell was right about one thing — the corruption of language. So I turned with interest to his article on the use of the French language in scientific communication. Abragam is a realist; he appreciates the position of English — "broken English" — and knows that nearly everyone will use it to make themselves widely understood. But while the Anglophones will suffer by seeing their language debased, the young French scientists, forced to learn a foreign language, will the better appreciate and use their own. Scientific French must however be preserved, and Abragam calls for a government subsidy to enable major books by French scientists to be published in French *and* English.

On big science and little science, in an article written in 1968, there is much good sense; we must have big science, he says, but must be careful not to lose our souls to it; "propter vitam, vitae perdere causas". And in an appendix written 15 years later for this volume, he wonders if big science and little science are terms that have lost their validity in this age of micro-electronics. He quotes with dismay the view of one of the most brilliant of American theorists, that the key to the unsolved problems of today lies in the computer. He hopes that in physics this is not so — and compares his feelings with those of the wife of Bishop Wilberforce, who hearing the doctrine that man is descended from the apes, said "if this horrible rumour turns out to be true, let us pray to God that it does not get widely known".

There is much else — a moving tribute to his friend and teacher John van Vleck; a letter on Orlov; and "La Physique, pour quoi Faire?", an address to the French Physical Society in 1973, which gives a passionate defence of our science. This last is perhaps a period piece; the call for relevance, the anti-intellectualism of popular culture, may have been stronger at that time than now. Physics stands for truth, he says, and the disorientation of youth is a result of the lies that pervade our society. He does not evade the question of scientists and defence. Yes, he says, to the scientists who worked on radar in the years before 1939, no, to those who were developing weapons for the Vietnam war. On the whole, he feels that scientists can have a good conscience if they do good science.

Professor Abragam has a lot to say; he says it with wit and clarity; and his French is a joy to read. □

Sir Nevill Mott was Cavendish Professor of Physics at the University of Cambridge from 1953 until his retirement in 1971. He was editor of the series International Monographs on Physics, *published by Oxford University Press, in which Professor Abragam's* Principles of Nuclear Magnetism *appeared in 1961.*

Insects organized

Ingrid H. Williams

Social Insects: Ecology and Behavioural Biology.
By M. V. Brian.
Chapman and Hall: 1983. Pp.377. Hbk £25, $49.95; pbk £12.95, $25.

THERE are few scientists with the breadth of knowledge and understanding of social insect biology who could attempt a comparative treatment of all eusocial insects. M. V. Brian is one of these few. In *Social Insects* he has succeeded in collating a vast amount of scattered research literature, most of it published during the last ten years, into a comprehensive treatise which will undoubtedly become a standard work of reference.

Evolutionary trends and comparative ethology are the underlying themes of the book. There are seventeen chapters, each of which is further subdivided to enable the reader to find particular topics without difficulty. Dr Brian starts with food collection and description of diverse foraging strategies of increasing complexity, moving on to give examples of social insect nest structure and construction, and an account of the means by which suitable microclimates are selected and maintained. A further chapter is devoted to the different weapons and social organizations evolved for effective colony defence.

Efficient food collection must be followed by efficient food processing, and here the importance of evolutionary trends to increase food preparation by adults for their young and to greater differentiation of a nursing caste to increase food conversion efficiency are stressed. This leads naturally to chapters on population growth, colony maturation and reproduction, with evaluation of the relative merits of the theories of kin selection, material manipulation and polygyne family selection in explaining the success of the evolution of sociality in Hymenoptera. Dr Brian next deals with maintenance of colony integrity, and compares the roles of queens in monogyne and polygyne colonies, before rounding off the book with a discussion of speciation and examples of the cooperative and competitive interrelationships involved in species coexistence at the community level.

The text is augmented by useful figures and some excellent photographs; the reference list is extensive and the subject and author indexes comprehensive. Written with the expert in mind, because it assumes a basic knowledge of social insects, non-specialists will also find the book fascinating and informative. □

Ingrid H. Williams is a Senior Scientific Officer in the Department of Entomology, Rothamsted Experimental Station, Hertfordshire.

© 1984 Nature Publishing Group